(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 173 254**
A1

(12) EUROPEAN PATENT APPLICATION

(21) Application number: 85110565.0

(22) Date of filing: 22.08.85

(51) Int. Cl.⁴: **C 12 N 15/00**, C 12 P 19/34, C 12 N 7/00, C 12 N 1/20, C 12 N 1/16, C 12 N 5/00, C 12 P 21/02, C 07 K 13/00, A 61 K 39/245, G 01 N 33/53, C 07 K 17/00 // (C12R1/19, 1:91)

(30) Priority: 23.08.84 EP 84110089
23.08.84 EP 84110090

(43) Date of publication of application: 05.03.86
Bulletin 86/10

(84) Designated Contracting States: AT BE CH DE FR GB IT LI LU NL SE

(71) Applicant: **Wolf, Hans Joachim,**
Josef-Jägerhuber-Strasse 9, D-8130 Starnberg (DE)

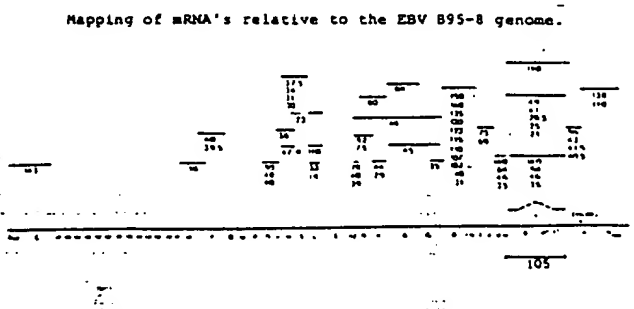
(72) Inventor: **Wolf, Hans Joachim,**
Josef-Jägerhuber-Strasse 9, D-8130 Starnberg (DE)

(74) Representative: **Vossius Vossius Tauchner Heunemann Rauh, Siebertstrasse 4 P.O. Box 86 07 67, D-8000 München 86 (DE)**

(54) DNA sequences of the EBV genome, recombinant DNA molecules, processes for producing EBV-related antigens, diagnostic compositions and pharmaceutical compositions containing said antigens.

(57) DNA sequences of the EBV genome, recombinant DNA molecules, processes for producing EBV-related antigens, diagnostic compositions and pharmaceutical compositions containing said antigens.

Described are DNA sequences of the EBV genome coding for EBV-related antigens, recombinant DNA molecules containing said DNA sequences, vector/host systems for cloning and expression of said DNA sequences, EBV-related antigens and methods for their preparation, diagnostic and pharmaceutical compositions containing said DNA sequences and antigens respectively (Fig. 2).



EP 0 173 254 A1

VOSSIUS · VOSSIUS · TAUCHNER · HEUNEMANN · RAUH
PATENTANWÄLTE

SIEBERTSTRASSE 4 · 8000 MÜNCHEN 86 · PHONE: (089) 47 40 75
CABLE: BENZOLPATENT MÜNCHEN · TELEX 5-29 453 VOPAT D

Hans Joachim Wolf
Josef Jägerhuber Str. 9
8130 Starnberg

August 22, 1985

DNA sequences of the EBV genome, recombinant DNA molecules, processes for producing EBV-related antigens, diagnostic compositions and pharmaceutical compositions containing said antigens

Technical field of invention

This invention relates to DNA sequences of the EBV genome coding at least for parts of EBV-related antigens to be used in methods and diagnostic and pharmaceutical compositions referred to below and methods of localising and isolating at least part of the respective DNA sequences.

Furthermore the invention relates to recombinant DNA - molecules i.e. cloning and expression vectors useful for the production of antigenic determinants of said EBV-related antigens after introduction of these vectors into appropriate hosts such as bacteria, yeasts and mammalian cells.

Finally this invention relates to methods and compositions or kits, respectively, for a rapid, simple, highly sensitive and highly specific determination of antibodies directed to EBV-related antigens. In these tests different antigens of EBV are used to detect specific antibody classes in the patient's serum, directed to these antigens. This detection allows fairly reliable conclusions as to the status of infection of the serum donor such as preinfection, fresh infection, chronic infection, convalescence and neoplastic condition. Furthermore, this invention relates to pharmaceutical compositions, e.g. vaccines containing said antigens useful for prophylaxis and therapy of EBV-related diseases.

Background Art

The herpesviruses (Herpetoviridiae) are enveloped icosahedral capsids with an overall diameter of 15o nm. The viral genome consists of a double-stranded DNA with a molecular weight of approximately $1o^8$ D. Human herpesviruses are Herpes simplex I ("fever blisters"), Herpes simplex II (genital herpes), Varicella-Zoster (chickenpox, shingles), Cytomegalovirus (congenital abnormalities, e.g. microcephaly), and Epstein-Barr virus (EBV) (infectious mononucleosis (IM), Burkitt's lymphoma (BL), nasopharyngeal carcinoma (NPC)

Herpesviruses display a remarkable propensity for establishing latent infections which may persist for the life of the host. After the primary infection the virus may remain quiescent, being demonstrable only sporadically or not at all, until it is reactivated by one of several known types of stimulus, such as irradiation or immunosuppression. Such exacerbations of endogenous disease may take the form of a crop of vesicles on the skin in the case of herpes simplex or zoster, or more generalized effects in the case of cytomegalovirus or EBV. The capacity to persist

indefinitely as a latent infection enables these viruses to survive in nature for a long time. During the last years, attention has turned to the correlation of human cancer and EBV.

Epstein-Barr-Virus (EBV), infections and their consequences

EBV causes infectious mononucleosis as a primary disease. Predominantly it affects children or young adults. More than 9o % of the average adult population is infected by EBV that persists for lifetime in peripheral B-lymphocytes. The virus is lifelong produced in the parotid gland and spread via the oral route.

Serology suggests that EBV might be involved in causing two neoplastic diseases of man, African Burkitt's lymphoma (BL) and nasopharyngeal carcinoma (NPC). Infectious mononucleosis is a consequence of primary infection by EBV. It is not a life-threatening disease if additional risk factors are absent.

However, the subjective feeling of sickness, frequently for extended periods (in the order of several weeks), and the necessity to avoid physical stress due to the drastically increased risk of spleenic rupture would certainly suggest a control of this disease.

The clinical diagnosis of infectious mononucleosis is usually derived from a combination of the following parameters:

1. High leukocyte count ranging from 1o,ooo to 2o,ooo and reaching up to 5o,ooo
2. 1o % atypical cells
3. lymphadenitis
4. fever.

Patients with infectious mononucleosis shed EBV in their saliva. Virus shedding does not require special prevention against spreading the disease as epidemics and infection of persons in close contact are rare (A.S. Evans, "The transmission of EB viral infections. Viral Infections in Oral Medicine.", edited by J. Hooks, G. Jordan, Elsevier North Holland Amsterdam, p. 211 (1982)). Virus shedding does not stop with recovery from disease and at least 60 % (possibly up to 100 %) of the adult population shed at least low levels of EBV which is produced lifelong in epithelial cells of the salivary duct of the parotid gland (H. Wolf, M. Haus, E. Wilmes, "Persistence of Epstein-Barr virus in the parotid gland", J. Virol. 51 (1984)).

About 1 % of the infectious mononucleosis cases show complications either already at the onset of the disease or as a late consequence. Most complications are due to autoimmune mechanisms and are in some cases indiscernable from graft versus host disease, a mechanism by which the body might clear itself from the excess of EBV converted proliferating B-cells.

If the T-cell response is insufficient, e.g. due to circumstances like treatment with high doses of Cyclosporin A in combination with corticosteroids or due to AIDS or a certain genetic predisposition as described by Purtilo (Duncan's syndrome, X-chromosome-linked lymphoproliferative disease (XLP); D.T. Purtilo, K. Sakamoto, V. Barnabei, J. Seeley, T. Bechtolg, G. Rogers, J. Yetz, S. Harada and the XLP-collaborators: "Epstein-Barr virus-induced diseases in boys with the X-linked lympho-proliferative syndrome (XLP). Update on studies of the registry." Am. J. Med. 73, p. 49 (1982)), infected B-cells may have a chance to escape from host control and grow without limitation as they would do when being

cultivated in vitro. The consequences have been described as BL-like disease in cases of AIDS patients (J.L. Ziegler, R.C. Miner, E. Rosenbaum, E.T. Lennette, E. Shillitoe, C. Casavant, W.L. Drew, L. Mintz, J. Gershor, J. Greenspan, J. Beckstead, K. Yamamoto, "Outbreak of Burkitt's-like lymphoma in homosexual men.", Lancet 2, p. 631 (1982)) or as a polyclonal lympho-proliferative disease for XLP-patients (D.T. Purtilo et al., supra) or kidney transplant recipients (D.W. Hanto, G. Frizzera, D.T. Purtilo, K. Sakamoto, J.L. Sullivan, A.K. Saemundsen, G. Klein, R.L. Simmons, J.S. Najarian, "Clinical spectrum of lympho-proliferative disorders in renal transplant recipients and evidence for the role of Epstein-Barr virus.", Cancer Res. 41, p. 4253 (1981)).

The positive and fast identification of infectious mononucleosis or acute EBV infection is especially important in cases where a differential diagnosis to leukemia or, in case of transplant recipient, to graft rejection crisis is necessary. In these cases, a false diagnosis may lead to incorrect therapy, which may have serious, even life-threatening effects.

Prevention of primary disease caused by EBV

Infectious mononucleosis seems to be unknown in areas like the Philippines or Malaysia (D.S.K. Tan, "Absence of infectious mononucleosis among Asians in Malaya.", Med. J. Malaya 21, p. 358 (1967)) where infection by EBV occurs very early in life. Almost the whole population has antibodies at the age of 2-10 years at the latest. Clinical symptoms seem to be a consequence of juvenile or adult infection. It can be assumed that a vaccine-primed organism will be infected without significant clinical symptoms and that the consequences often fatal in the risk groups listed above could be eliminated by a vaccine.

Burkitt's Lymphoma and EBV

The development of Burkitt's lymphoma is linked to chromosomal rearrangements. Not all cases contain EBV genomes in the tumor cells. However, at least in areas with high incidence, 97 % of these neoplasias are EBV-related and a control of EBV infection is likely to reduce the risk of developing Burkitt's lymphoma.

Nasopharyngeal carcinoma as a possible "secondary disease" related to EBV

The other disease where EBV shows a 100 % association is nasopharyngeal carcinoma (NPC) ("The Biology of Nasopharyngeal Carcinoma", UICC technical report series, vol. 71, edited by M.J. Simons and K. Shanmugaratnam, International Union Against Cancer, Geneva, p. 1 (1982)). NPC most frequently starts at the fossa of Rosenmueller (Recessus pharyngeus) at the postnasal space. Frequently patients are hospitalized only after the first typical metastases have developed in the cervical lymph nodes.

In some areas of Southern China and amongst Chinese in Singapore and Malaysia, NPC is the most frequent neoplasia of man with an incidence of up to 40 per 100,000 per year. In other parts of the world, like Borneo or Tunesia the incidence is also high. In most other areas, the incidence is around 0,2 per 100,000 per year which represents about 4 % of ear, nose and throat (ENT)-tumors. The age distribution shows a clear single peak around the age of 40 to 50 in almost all high-risk areas. In Borneo and to some extent in Tunesia, a remarkable second peak has, however, been observed at an early age ranging from 5 to 15 years (M. J. Simons et al., supra).

Environmental factors including traditional Chinese medicine may be responsible for the increased risk of nasopharyngeal carcinoma in certain, predominantly Chinese, populations of Southern Asia (H. Wolf, "Biology of Epstein-Barr virus in: "Immune deficiency and cancer: Epstein-Barr virus and lymphoproliferative malignancies", ed. D. Purtilo, Plenum Press, p. 233 (1984)).

Control of EBV-related neoplasia

There are three possible basic strategies to control neoplasia:

1. Early detection followed by therapy,
2. delay of onset of disease ideally beyond the average lifespan, and
3. prevention.

These goals may be achieved also in multifactorial diseases such as many neoplasias. Incidence of disease may be reduced by eliminating one or more of the essential factors which are not necessarily sufficient by themselves to cause the disease, or by reducing factors which promote the manifestation of neoplastic conditions. The use of the specific virus-related antigens of this invention, or antibodies or genetic materials as tools for early diagnosis of virus-related tumors, might facilitate the elimination of essential factors.

Selection of EBV-related gene products for diagnosis of EBV-related NPC

A. Primary infection with EBV: Development of antibodies against VCA (viral capsid antigen), EA (early antigen) and EBNA (Epstein-Barr Nuclear Antigen)

EBV infects B-lymphocytes during acute or primary infection (mononucleosis). Due to the lack of immune response, a number of cells enter into the lytic cycle and produce a full set of viral antigens which are shed into the blood stream during cytolysis. Against these antigens, specific antibodies will be synthesized by the host's immune system (Table A).

Probably not all B-lymphocytes are capable of supporting a fully lytic infection due to a cellular factor which prevents expression of EBV. These cells are latently carrying EBV genomes for the rest of the host's life.

Table A

DISEASE	VCA: IgG	IgM	IgA	EA	EBNA	MA[1]
NORMAL ADULTS	+	-	-	-	+	+
ACUTE ADULTS (EARLY)	++	+	-	+	-	-
CHRONIC INFECTION	+	+	-	±	±	±?
REACTIVATION	+	+	-	+	+	+
XLP[2]	+	-	-	±	(+)	?
NPC	++	-	+	+(D)	+	++
BL	++	-	-	+(R)	+	+

[2] XLP AS AN EXAMPLE OF IMMUNOLOGICALLY DEPRIVED HOSTS

[1] DETERMINED BY IMMUNOPRECIPITATION OF GP 240/200 (MA; membrane antigen)

B. Convalescence: Disappearance of antibodies against EA and maintenance of antibodies against VCA and EBNA

As the immune defense mechanisms of the body remove the

lytically infected cells from the circulation, the antibody levels will start to fall during the convalescent phase. After a certain period, anti-EA-antibodies disappear. However, as mentioned above, EBV is produced in the parotid gland. The viral particles and intracellular virus-associated antigens including EA will be shed into the saliva and reach the oropharynx. Here the viral particles bind to the B-lymphocytes and are presented to the body as antigens, thus the antibody titer against VCA is maintained. Since EA cannot bind to the lymphocytes it will be degraded by poteases and therefore will not be available to the immunesystem as an antibody-inducing antigen.

The circulating lymphocytes that are latently infected by EBV contain EBNA. At the end of their life cycle these cells disintegrate and release EBNA into the blood stream. Therefore antibodies to this antigen will persist.

Thus, due to the EBV-production in the parotid gland and to the release of EBNA from latently infected B-cells, sera of convalescents will have low anti-VCA and anti-EBNA IgG-antibody levels (see Table A, supra). In addition EA released from rare B lymphocytes which may enter a lytic cycle may be an inferior antigen and may not give rise to antibody levels detectable with the test systems used.

In combination with the known sequence of appearance of antibody classes, specifically the early presence of IgM antibodies followed by IgG antibodies, the various antigen classes of primary disease caused by EBV can be utilized for improved diagnostic procedures. However, available test systems which are mainly based on cellular antigens or cell derived antigens have serious limitations. This concerns the sensitivity, especially for detection of IgM antibodies and also unspecific reactions.

C. EBV-related antibodies in individuals suffering of NPC:

The first suggestive evidence that Epstein-Barr virus might be causally related to nasopharyngeal carcinoma and African Burkitt's Lymphoma was derived from serological data (for review see M.A. Epstein, B.G. Achong, "The Epstein-Barr Virus" Springer Verlag Berlin, Heidelberg, New York (1979)).

Using mainly indirect immunofluorescence on cells producing virus or at least early viral antigens, significantly higher antibody titers to these antigens were found in patients' sera. These first tests which detected unspecified immunoglobulin classes against a group of proteins named Early Antigen (EA) and another group of proteins named Virus Capsid Antigens (VCA) were helpful for the establishment of a relationship between EBV and these diseases. These tests, however, are of limited value for definite diagnosis of the malignancies from a single serum, and cannot be used for monitoring therapy.

The introduction of antigen and antibody class specific tests, specifically the determination of peripheral IgA antibodies for the two antigen families EA and VCA and also the first attempts to subdivide at least the EA-family (EA, D or R; G. Henle, W. Henle and G. Klein, "Demonstration of 2 distinct components in the early antigen complex of Epstein-Barr Virus infected cells", Int. J. Cancer 8, p. 272 (1971)) achieved remarkable improvements of the diagnostic and prognostic value of the tests.

In the areas of high risk for NPC, 1% of the adult population has IgA antibodies for EBV-Capsid antigen (VCA).

Three percent of this group has NPC upon clinical examination and, with the exception of terminal cases, there were no anti-VCA IgA negative cases detected. Out of the IgA anti-VCA positives, about 1% per year developed NPC in a 3 year follow up. A test of this quality, if available as a highly specific automat- readable ELISA test, would provide an excellent "first step" screening for a population of extreme risk.

Detection of EB virus IgA/VCA antibody is helpful for diagnosis of NPC (see table on page 14), and of special value for the detection of early stages. For example, in Wuzhou City (China; high risk area for NPC), the frequency of NPC detected by serological mass survey revealed a much higher percentage of patients in stages I (42%) and II (48%) than otherwise detected in outpatient clinics (1.7% stage I and 3o % stage II). The chance of survival is clearly related to the stage at which therapy is begun. The survival rates for stage I are (according to Shanghai Tumor Hospital) 93%, for stage II 75%, and are very low for more advanced stages. Therefore it is possible to reduce the mortality rate of NPC through early detection and early treatment.

IgA antibodies to the early antigen complex of EBV can be detected in 40 % to 70 % of NPC patients, depending on the method used. These antibodies are virtually absent in the non-tumorbearing population. Such test of the tumorbearing individuals should be of great importance for the decision to start therapy, and its value would be even higher if the sensitivity could be enhanced to allow detection of disease in closer to 100 % of the tumor patients.

The detection rate of NPC among IgA/VCA antibody-positive individuals is 1.9 % and that of IgA/EA individuals is 30-40 %. These data indicate that the IgA/EA antibody

test is more specific for the detection of NPC, but not as sensitive as IgA/VCA antibody.

A number of laboratories have used the continuous determination of IgA antibodies to EA and VCA to monitor the success of therapy and for early detection of relapse with very good success.

Membrane protein gp 250/350 and its use

Four proteins of the viral envelope constituting the so-called membrane antigen complex (MA) have been described (L.F. Qualtiere, G.R. Pearson, et. al., supra; J. North, A.J. Morgan, M.A. Epstein, "Observations on the EB virus envelope and virus-determined membrane antigen (MA) polypeptides", Int. J. Cancer 26, p. 231 (1980)). Two of these proteins, i.e. gp 250 and gp 350, are antigenically closely related (D.A. Thorley-Lawson and K. Geilinger, "Monoclonal antibodies against the major glycoprotein (gp 250/350) of Epstein-Barr virus neutralize infectivity", Proc. Natl. Acad. Sci, USA 77, p. 5307 (1980)). The molecular weight of one component ranges from 200,000 to 250,000 D depending on the cell line where the virus is derived from and the second antigenetically related glycoprotein has a molecular weight of 300,000-350,000 D but is absent in some cell lines. Since these glycoproteins are all related in antigenicity, protein and encoding DNA sequence, they are usually referred to as gp 220/350 or gp 250/350 or simply as gp 250 or gp 350 but meaning the whole family of related glycoproteins.

Glycoprotein 250/350 is able to bind to the EBV receptor of human and some primate B-lymphocytes and to thus initiate the infection of these cells (A. Wells, N. Koide, G. Klein, "Two large virion envelope glycoproteins mediate Epstein-Barr virus binding to receptor-positive cells",

J. Virol. 41, p. 286 (1982)). Antibodies against these proteins neutralize the infectivity of the virus, which could be demonstrated for human as well as for rabbit antisera and mouse monoclonal antibodies (D.A. Thorley-Lawson et. al., supra). By the use of monoclonal antibodies it has been shown that blocking of only one antigenic determinant present both in gp 350 and gp 250 was sufficient for virus neutralization. Adsorption of human sera to immobilized gp 350 and gp 250 removed the neutralizing antibodies (D.A. Thorley-Lawson et. al., supra). Thus, there is convincing evidence that

a) gp 350 and gp 250 induce the production of neutralizing antibodies, and that
b) antibodies against gp 350 and gp 250 have neutralizing capacity.

Therefore, this protein as well as its related viral gene product, gp 350 (with a molecular weight of 350,000), are candidates for a possible EBV vaccine (A.J. Morgan, M.A. Epstein, J.R. North, "Comparative immunogenicity studies on Epstein-Barr virus membrane antigen (MA) gp 34o with novel adjuvants in mice, rabbits and cotton-top tamarins", J. Med. Virol. 13, p. 281 (1984)). These glycoproteins are expressed on induced EBV producer cell lines and can be easily demonstrated after radioiodination of cell surface proteins (L.F. Qualtiere, G.R. Pearson, "Epstein-Barr virus-induced membrane antigens: immunochemical characterization of Triton X-1oo solubilized viral membrane antigens from EBV superinfected Raji cells", Int. J. Cancer 23, p. 808 (1979)).

Application of gp 25o/35o for the diagnosis of EBV-related diseases

IgG antibodies are absent during the acute phase of primary EBV infection, but present for lifetime after convalescence. IgM antibodies are present in the early stage of the disease and absent during convalescence.

IgA antibodies against EBV-antigens are present almost exclusively in NPC patients and can be detected in sera of at least 58 % of these patients even with not very sensitive tests (Zeng Yi and Hans Wolf, manuscript in preparation and example 16, infra).

Comparison of Positive Rate of IgG and IgA Antibodies to VCA and MA from NPC Patients and Normal Individuals

Cases		MA/IgG		MA/IgA		VCA/IgA		EA/IgA	
		(+) No.	(+) rate%	(+) No.	(+) rate%	(+) No.	(+) rate%	(+) No.	(+) rate%
NPC Patients	48	48	100	28	58.3	48	100	31	64.6
Normal Individuals	48	47	97.9	0	0	0	0	0	0

* MA/IgG and MA/IgA detected by immunofluorescence test

VCA/IgA and EA/IgA detected by immunoenzymatic test

The whole gp 25o molecule or parts of its backbone polypeptide chain can be utilized as reagents in preferentially class-specific antibody detection tests such as passive hemagglutination, counter gel electrophoresis, radioimmunoassays or enzyme-linked immuno-absorbent assays.

Highly specific test antigens allow better signals and detect otherwise unrevealed low antibody levels of clinical significance. The use of singular antigenic sites of the gp 25o instead of the entire gene product may, in some cases, permit a more precise diagnosis of the disease.

Application of gp 25o/35o for prophylaxis and treatment of EBV-related diseases

A. Since infection by EBV early in life only causes subclinical seroconversion, it may be anticipated that the presence of maternal antibodies or antibodies induced by a vaccine will influence the clinical manifestation of a primary EBV infection. It is expected that the vaccination of children or young adults, preferably before the peak of risk of catching an EBV infection, reduces effectively the clinical manifestation of infectious mononucleosis in the population.

B. In all areas with high incidence rates for NPC or BL, the population shows almost 1oo % seroconversion to EBV within the first one to two years of life. Vaccination will have to take place soon after birth. If this vaccination is regularly repeated, it will in all probability prevent EBV infection, delay it or reduce the biological effects of early primary infection. Each of these consequences is expected to either prevent the subsequent development of neoplasia, to delay its onset considerably or to decrease the relative risk.

C. In NPC, occasional production of viral antigens at the site of the tumor will stimulate primarily IgA secreting B-lymphocytes. IgA antibodies are capable of blocking antibody-mediated cytotoxicity. IgA antibodies to viral membrane antigens, such as gp 25o, are present in NPC and BL patients and may not only be indicators of the disease, but may even contribute to the failure of the immune system to eliminate the tumor cells by their masking potential. Large doses of the purified antigen given to tumor patients may bind IgA and initiate the formation of an excess of IgG antibodies directed to the same antigen. These specific IgG antibodies may then compete with remaining IgA antibodies and allow the elimination of tumor cells by antibody-dependent mechanisms.

D. Appropriate administration of gp 25o or related products might also enhance the cellular immune mechanisms and thus restrict the growth of tumors.

Production of EBV specific antigens according to the present invention

1. As a consequence of all findings, it is one of the objects of this invention to improve the sensitivity of tests for detection of antibody classes and antigen specific antibodies and to develop a system which allows mass testing and better standardization.

2. EBV cannot be efficiently produced in a lytic cell cycle since efficiently infectable cells are not known at present and because all of the cells used as source for the preparation of EBV or related antigens are immortalized cells or even tumor derived cells. In most cell lines retroviruses have been demonstrated. The products isolated from such cultures therefore are not only very expensive but their use is also a potential safety risk.

3. The application of recombinant DNA technology has made possible the production of useful polypeptides by appropriate host cells transformed with recombinant DNA molecules and grown in appropriate culture systems.

4. According to the present invention, recombinant DNA methods are used to express the genetic information of the genes or at least of parts of the genes encoding the EBV proteins p138, p150 and gp 250/350 in appropriate host cells, such as bacteria (e.g. the genera Escherichia, Salmonella, Pseudomonas or Bacillus), yeasts (e.g. the genera Candida, or Saccharomyces) and mammalian cells (e.g. Vero-cells, CHO-cells or lymphoblastoid cell lines).

5. Furthermore, the genomic regions encoding the EBV proteins p150, p143, p138, p110, p105, p90, p80 and p54 were identified and their relevance for diagnostic purpose has been identified. Therefore, the key information for the production of these proteins or antigenic determinants thereof in a manner as demonstrated for the proteins p138, p150 and gp 250/350 is also disclosed in the present invention.

Recombinant DNA technology

A. Expression control systems

Prokaryotes most frequently are represented by various strains of E. coli. However, other microbial strains may also be used, such as bacilli, for example Bacillus subtilis, various species of Pseudomonas, or other bacterial strains. In such prokaryotic systems, plasmid vectors which contain replication sites and control sequences derived from a species compatible with the host are used. For example, E. coli is typically transformed using derivatives of pBR322, a plasmid derived from an E. coli species by Bolivar, et al., Gene 2, p. 95 (1977). pBR322 contains genes for

ampicillin and tetracycline resistance, and these markers can be either retained or destroyed in constructing the desired vector. Commonly used prokaryotic control sequences which are defined herein to include promoters for transcription initation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (lac) promoter systems (Chang, et. al. Nature 198, p. 1056 (1977)) and the tryptophan (trp) promoter system (Goeddel, et al., Nucleic Acids Res. 8, p. 4057 (1980)). The lambda-derived P_L promoter and N-gene ribosome binding site (Shimatake, et al., Nature 292, p. 128 (1981), which has been made useful as a portable control cassette are further examples. However, any available promoter system compatible with prokaryotes can be used.

In addition to bacteria, eukaryotic microbes, such as yeast, may also be used as hosts. Laboratory strains of Saccharomyces cerevisiae, Baker's yeast, are most used, although a number of other strains are commonly available. While vectors employing the 2 micron origin of replication are illustrated (J.R. Broach, Meth. Enz. 101, p. 307 (1983)), other plasmid vectors suitable for yeast expression are known (see, for example, Stinchcomb et al., Nature 282, p. 39 (1979); Tschempe et al., Gene 10, p. 157 (1980) and L. Clarke et al., Meth. Enz. 101, p. 300 (1983)). Control sequences for yeast vectors include promoters for the synthesis of glycolytic enzymes (Hess et. al., J. Adv. Enzyme Reg. 7, p. 149 (1968); Holland et al., Biochemistry 17, p. 4900 (1978)). Additional promoters known in the art include the promoter for 3-phosphoglycerate kinase (Hitzeman et al., J. Biol. Chem. 255, p. 2073 (1980)), and those for other glycolytic enzymes, such as glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate iso-

merase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. Other promoters, which have the additional advantage of transcription controlled by growth conditions are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and enzymes responsible for maltose and galactose utilization (Holland, supra).

Evidence suggests that terminator sequences are desirable at the 3' end of the coding sequences. Such terminators are found in the 3' untranslated region following the coding sequences in yeast-derived genes. Many of the vectors illustrated contain control sequences derived from the enolase-I gene containing plasmid peno46 (M.J. Holland et al., J. Biol. Chem. 256, p. 1385 (1981)) or the LEU 2 gene obtained from YEp13 (J. Broach et al., Gene 8, p. 121 (1979)), however, any vector containing a yeast-compatible promoter, origin of replication and other control sequences is suitable.

It is also, of course, possible to express genes encoding polypeptides in eukaryotic host cell cultures derived from multicellular organisms. See, for example, Cruz and Patterson, editors, "Tissue Cultures", Academic Press (1973). Useful host cell lines include VERO and HeLa cells, and Chinese hamster ovary (CHO) cells. Expression vectors for such cells ordinarily include promoters and control sequences compatible with mammalian cells such as, for example, the commonly used early and late promoters from Simian Virus 40 (SV 40) (Fiers et al., Nature 273, p. 113 (1978)), or other viral promoters such as those derived from polyoma. Adenovirus 2, bovine papiloma virus, or avian sarcoma virusses. General aspects of mammalian cell host system transfor-

mations have been described by Axel in U.S. Patent No. 4,399,216 issued August 16, 1983. It now appears also that "enhancer" regions are important in optimizing expression; these are, generally, sequences found frequently upstream of the promoter region. Origins of replication may be obtained, if needed, from viral sources. However, gene integration into the chromosome is a common mechanism for DNA replication in eukaryotes, and hence independently replicating vectors are not required. Plant cells are also now available as hosts, and control sequences compatible with plant cells such as the nopaline synthase promoter and polyadenylation signal sequences (A. Depicker et al., J. Mol. Appl. Gen. 1, p. 561 (1982)) are available.

B. Transformation of suitable hosts

Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described by S.N. Cohen, Proc. Natl. Acad. Sci. (USA) 69, p. 2110 (1972) is used for prokaryotes or other cells which contain substantial cell wall barriers. Infection with Agrobacterium tumefaciens (C.H. Shaw et al., Gene 23, p. 315 (1983)) is used for certain plant cells. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb (Virology 52, p. 546 (1978)) is preferred. Transformations into yeast are carried out according to the method of P. van Solingen et al. (J. Bact. 130, p. 946 (1977)) and C. L. Hsiao et al. (Proc. Natl. Acad. Sci. (USA) 76, p. 3829 (1979)). Alternatively, the procedure of Kiebe, et al. (Gene 25, p. 333 (1983)) can be used.

C. Construction of recombinant cloning and expression vectors

Construction of suitable vectors containing the desired coding and control sequences employs standard ligation and restriction techniques which are well understood in the art. Isolated plasmids, DNA sequences, or synthesized oligodeoxyribonucleotides are cleaved, tailored and religated in the form desired.

Site specific DNA cleavage is performed by treating with the suitable restriction enzyme (or enzymes) under conditions which are generally understood in the art, and the particulars of which are specified by the manufacturer of these commercially available restriction enzymes. See, e.g., New England Biolabs, Product Catalog.

If desired, size separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. A general description of size separations is found in "Methods in Enzymology" 65, p. 499-560 (1980).

Restriction cleaved fragments may be blunt ended by treating with the large fragment of E. coli DNA polymerase I (Klenow) in the presence of the four deoxynucleotide triphosphates (dNTPs). The Klenow fragment fills in at 5' sticky ends but chews back protruding 3' single strands, even though the four dNTPs are present. If desired, selective repair can be performed by supplying only a selected one or more dNTPs within the limitations dictated by the nature of the sticky ends. Treatment under appropriate conditions with S1 nuclease results in hydrolysis of any single-stranded portion.

Synthetic oligonucleotides may be prepared by the triester method of Matteucci et al. (J. Am. Chem. Soc. 103, p. 3185 (1981)) or the diethylphosphoramidite method of Caruthers, described in U.S. Patent No. 4,415,732, issued November 15, 1983.

Ligations are performed under standard conditions and temperatures (as described below) using T4 DNA ligase. In vector constructions employing "vector fragments", the vector fragment is commonly treated with bacterial alkaline phosphatase (BAP) in order to remove the 5' phosphate and prevent religation of the vector. BAP digestions are carried out under standard conditions (as described below).

D. Selection of transformants

In the constructions correct ligations for plasmid construction are confirmed by transforming E. coli or other suitable hosts with the ligation mixture. Successful transformants are selected by ampicillin, tetracycline or other antibiotic resistance or using other markers depending on the mode of plasmid construction, as is understood in the art.

Brief summary of the invention

The present invention relates to the production of EBV specific antigens by recombinant DNA technology and their use in diagnosis, prophylaxis and therapy of EBV-related dieseases. Therefore, it is an object of this invention to identify novel Epstein-Barr viral antigens, such as p150, p143, p138, p110, p105, p90, p80, p54 (G.J. Bayliss, H. Wolf, infra), which are correlated with Epstein-Barr virus related diseases like nasopharyngeal carcinoma (NPC), infectious mononucleosis, and Burkitt's lymphoma (see legend to Fig. 1 and Fig. 28) by immunological methods.

Another object of this invention is the localization and identification of genomic regions of EBV, for example as it has been cloned from B95-8 cells (American Type Culture Collection, Rockville, Maryland, USA (ATCC) CRL1612) (J. Skare, J.L. Strominger, "Cloning and mapping of BamHI endonuclease fragments of the DNA from the transforming B95-8 strain of Epstein-Barr Virus", Proc. Natl. Acad. Sci. USA 77, p3860(1980)) coding for said antigens of diagnostic importance and of relevance for medical purposes. This is achieved by using the hybrid selection method.

A further object of the present invention is the subcloning of a genomic region of EBV, for example from existing libraries of EBV, cloned from B95-8 cells which encodes at least a part of useful antigens for medical purposes, such as p138 and p150. This is achieved by joining a subgenomic fragment, e.g. an XhoI-fragment, derived from the EBV B95-8 subclone pBR322 BamA, (J.Skare et al., supra) to the plasmid pUC8 (J. Messing, infra) (pUC635, see Fig. 4).

Another object of this invention is the production of proteins by expression of the respective genetic information in suitable host cells, such as bacteria (e.g. of the genera Escherichia, Salmonella, Pseudomonas or Bacillus), yeasts (e.g. of the genera Candida or Saccharomyces) animal cells and human cells (e.g. Vero-cells; CHO-cells; CHO dhfr$^-$ cells in combination with an appropriate selection system, optionally a plasmid carrying a functional dhfr gene as well as the genetic information for the EBV gene under control of a suitable regulation sequence; or lymphoblastoid cell lines). The proteins produced by these host cells contain e.g. p138, p150 or gp250/350-related antigenic determinants (epitopes) and are, depending on the expression system, synthesized either as a fusion protein or as a non-fusion protein.

For the production of a fusion protein by bacteria the expression of the genomic subfragments, for example that encoding a part of p138 of EBV B95-8 and introduced into the known plasmid pUC8, was induced e.g. by isopropyl-ß-D-thioglactopyranoside (IPTG). The respective expression products were identified by immunological methods.

Another fusion protein is provided by cleaving subclone pUC635 with EcoRI and BglII and introducing this fragment into the vector plasmid pUC9 (U. Rüther, infra). The resulting recombinant plasmid is pUC924 (Fig. 6). The expression product has a size of about 94 kd.

A further fusion protein is produced by expressing the genetic information of said XhoI-p138-encoding fragment of pBR322 BamA in the plasmid pEA305 (E. Amann, J. Brosius, M. Ptashne, "Vectors bearing a hybrid trp-lac-promoter useful for regulated expression of cloned genes in E. coli", Gene 25, p. 167 (1983)). After putting the p138 related information into a proper reading frame relative to pEA305, the clone pMF924 synthezises a fusion protein that contains a part of the λ-repressor protein c_1 (Fig. 7).

Still another fusion protein is provided by cloning a 3.0 kb genomic XhoI-fragment containing p138-related genetic information 3' to a hybrid trp-lac promoter (as described by F. Amann et. al., supra). For this purpose the known plasmid pKK240-11 was used. The resulting clone pKK378 synthezises a fusion protein that is composed of an aminoterminal methionine residue followed by p138 related DNA sequences (Fig. 8).

Still another object of the present invention is to provide fusion proteins, non-fusion proteins or oligopeptides which contain only the antigenic determinant protein subregions of viral proteins like p138. For this purpose the determinants of the protein are located by a computer-directed analysis, using a computer program developed by

us for the Digital Equipment VAX 11/750 computer. A similar program has been used for other problems and another computer by G.H. Cohen, B. Dietzschold, M. Ponce de Leon, D. Long, E. Golub, A. Varrichio. L. Pereira, R.J. Eisenberg, "Localization and synthesis of an antigenic determinant of Herpes simplex virus glycoprotein D that stimulates the production of neutralizing antibody", J. Virol 49, p. 102 (1984). By cloning the respective fragments into vectors like pUC8 or pUR288 (U. Rüther et al., infra) plasmids as pUR600 and pUR540 were obtained. The produced large and small fusion proteins are investigated by gel electrophoresis and immunoblotting experiments. The cloning experiments in pUR288 were done for stabilizing the small p138-related polypeptides.

A further object of the present invention is the expression of polyantigens composed of antigenic determinants of several different EBV-serotypes. For that purpose the corresponding DNA fragments are linked and introduced in a suitable vector. The expression products are fusion and non-fusion EBV-specific polyantigens.

Further fusion proteins containing p150-related antigenic determinants were obtained by cloning and expression of the corresponding DNA sequences in pUR plasmids and pUC plasmids. The obtained constructs were the recombinant plasmids pUR290CXH580, pUR290DBX320, pUR292DBB180, pUR290DTT700, pURDTT740, pUR290DTP680, and pUR288DPP320.

Another object of the present invention is the construction of new expression vectors, such as pUC600 and pUC601 which contain a part of the coding region of the viral protein p138. If DNA sequences are introduced 3' to this sequence into the vector, the expression products are stabilized by the p138-specific aminoacid sequence and protected against protease degradation.

Still another object of this invention is the modification of said expression vectors by introducing a DNA-sequence coding for three to fifteen Arginine residues and at least one stop codon 3' of the cloning site of said expression vectors and furthermore positioned in an appropriate reading frame. The obtained vector is pUCARG601. If DNA-sequences coding for proteinaceous material are inserted into this expression vector the expression products will be fusion proteins carrying said Arginine residues at their carboxy terminus, such as those fusion proteins encoded by plasmids pUCARG1140 (see Fig. 12a)) and pUCARG680.

Thus it is an object of this invention to provide a simple method for isolating proteins useful for diagnosis, prophylaxis and therapy such as EBV p138 or related polypeptides or oligopeptides (antigens) from the host cell lysate according to the method of H. M. Sassenfeld, S.J. Brewer ("A polypeptide fusion designed for the purification of recombinant proteins", Bio/Technology 2, p. 76 (1984)).

By introduction of said Arginine residues the net charge of the expressed proteins becomes more positive and after lysis of the host cells the oligo-arginine linked proteins are isolated by a SP Sephadex C-25 column chromatography. Due to the oligo-arginine group the EBV specific proteins are eluted at a high NaCl-concentration. This eluate is then treated with carboxypeptidase B which degrades carboxy-terminal lysine and arginine residues. Finally another SP Sephadex C-25 chromatography is carried out wherein the EBV-related proteins are eluted at low salt concentrations (see Fig.16). It is evident, however, that this procedure may be used also for the purification of proteins secreted into the medium.

It is also evident that other established methods for protein-purification such as molecular sieving or affinity chromatography on ion exchange columns or columns loaded with specific antibodies to the expressed proteins can be used as additional or alternate purification methods. For the production of non-fusion proteins which essentially contain amino acid sequences of the naturally occurring proteins or parts thereof the recombinant plasmids of the present invention may be modified. If an oligonucleotide linker is inserted between the bacterial protein encoding region and the EBV-related protein encoding region of the expression vector, the amino acid sequence corresponding to the oligonucleotide linker becomes part ot the expressed fusion protein. After isolation of this fusion protein from the transformants expressing it, it is cleaved either by amino acid sequence specific proteases in the introduced aminoacid linker or, if the amino acid linker comprises peptide bonds sensitive to acid cleavage, by treatment with acids, e.g. formic acid.

A further object of this invention is the cloning of a genomic region of EBV coding for at least a part of the specific viral antigen gp 250 and gp 350. This is achieved by joining a subgenomic PstI-PstI fragment of the EBV genome from the cell strain 395-8 (ATCC CRL 1612) (R. Baer et al., infra) contained in pBR322 BamL (J. Skare et al, supra) to the plasmid pUC8 (J. Messing et al., infra). The resulting recombinant plasmid is designated as pUCLP1.9 (see Fig. 19).

For the production of a fusion protein by bacteria, a genomic subfragment coding for a part of gp 250 and gp 350 of EBV B95-8 was cloned into the vector pUR 290 (U. Rüther et al., infra) which carries a region of the lacZ gene coding for the enzyme ß-galactosidase (pURLP1.9, see Fig. 20). The respective expression product was purified and identified by immunological methods.

Still another object of the present invention is to provide fusion proteins or non-fusion proteins which contain only the antigenic determinant protein subregions of gp 250 and gp 350. For this purpose the antigenic determinants of the proteins were localized by a computer-directed analysis using a computer program developed by us for the Digital Equipment VAX 11/750 computer.

The respective DNA-fragments are then cloned in a conventional expression vector such as pUR (ß-galactosidase) (U. Rüther, et al., infra). Plasmids obtained were e.g. pURLEP600 and pURLXP390 (see Fig. 27).
Furthermore, the N-terminal antigenic determinant of gp250/350 was expressed as a fusion protein in a pUC vector (pUCLEP600, see Fig. 27).
Another fusion protein is provided by cloning a DNA-fragment coding for the N-terminal antigenic determinant of gp 250 and gp 350 into the expression vector pUCARG601 mentioned above.

A further object of the present invention is the expression of polyantigens containing several antigenic determinants of gp 250 and gp 350 located by said computer analysis. For this purpose the corresponding DNA fragments are linked and introduced in a suitable vector. The expression products are fusion and non-fusion EBV-specific polyantigens.

A final object of the present invention is the utilization of either said EBV-related proteins or subregions thereof or, if suitable, EBV-related DNA fragments or clones, for the production of diagnostic compositions (kits) useful in clinical diagnosis or scientific research. These tests are based on principles as ELISA (Enzyme-linked immuno sorbent assay), RIA (Radio immuno assay) or the indirect hemagglutination assay. Furthermore, the EBV-

related proteins can be used, e.g. for monitoring vaccination programs, analyzing epidemiological problems, for patients treatment, and for the production of vaccines for prophylaxis and therapy of EBV-related diseases, such as mononucleosis, Burkitt's lymphoma and nasopharyngeal carcinoma. Vaccines can be manufactured according to conventional methods. Unit doses are filled in vials optionally together with a conventional adjuvant such as aluminium hydroxide. Alternatively the product may be administered in the form of aggregates with liposomes. Patients may be vaccinated with a dose sufficient to stimulate antibody formation and revaccinated after one month and after 6 months.

Finally the proteins are useful for prophylaxis and therapy of EBV-related diseases, because they are able to modulate the immune response in patients suffering from diseases such as NPC, chronic infectious mononucleosis or EBV-related Burkitt's lymphoma.

Brief description of the drawings

Figure 1: Autoradiography of an immunoprecipitation of EBV-specific sera derived from patients suffering from mononucleosis and NPC.

The immunoprecipitated ^{35}S-labelled proteins were separated by a SDS-polyacrylamide gel electrophoresis and an X-ray-film was exposed to the gel.

The sources of the different sera used for precipitation are given at the bottom of the respective regions of the autoradiography. The control, designated "pool", contains all of the immunoprecipitatable EBV-specific proteins.

It can be taken from the autoradiography that at least antibodies to p138, p105 and p80 are present in each of the NPC sera and only in some of the other EBV-infection specific sera. In analogy, antibodies to p54 are significant for fresh EBV infection (infectious mononucleosis) as compared to convalescent state. Antibodies to p150, p143, p110, p90 are also present in convalescent sera of healthy individuals and can serve as markers for immunity or, in connection with IgM specific tests, for fresh EBV infection or, in connection with IgA, for a specific test for EBV-related neoplasia (NPC and BL).

Figure 2: Mapping of mRNA's relative to the EBV B95-8 genome.

The BamHI restriction sites of the EBV B95-8 genome are given at the bottom of the figure and the respective restriction fragments are designated by upper and lower case letters. The mRNA's of the proteins localized by hybrid-selection to individual BamHI restriction fragments are indicated by numbers and lines.
It can be taken from the figure, that the gene of p138 was correlated to the BamA-fragment.

Figure 3: DNA sequence of the leftward reading frame of BamA encoding p138.

The sequence shown is the respective negative strand. The p138 encoding region startsat nucleotide position 182 and ends at nucleotide position 3565. Restriction sites used for cloning of fragments of this coding region are indicated.

Figure 4: Restriction map of the plasmids pUC635 and pUC6130.

The size of the vector pUC8 is 2.7 kb. The cloning site 3' of the lacUV5 ß-galactosidase promoter and operator (PO) contains EcoRI(E), BamHI(B), SalI(S), PstI(P), and HindIII(H) site. The ß-lactamase gene is indicated by AMP. The 3.0 kb and 3.3 kb XhoI-fragments of the p138 coding region are inserted into the SalI site of pUC8. The insertion is indicated by an open bar. pUC635 contains the 3.0 kb XhoI-fragment in a correct reading frame relative to the ß-galactosidase gene, whereas pUC6130 contains the 3.3 kb XhoI-fragment in the opposite orientation.

Figure 5: Expression of the proteins encoded by plasmids pUC635, pUC924, pMF924, and pKK378.

Lane 1 : of the immunostained Western-blot shows the proteins isolated from bacteria transformed with pUC8 and induced with IPTG.

Lane 2: proteins of pUC924 transformed bacteria
Lane 3: proteins of pKK378 transformed bacteria
Lane 4: proteins of pMF924 transformed bacteria
Lane 5: proteins of pUC635 transformed bacteria
The size of the fusion protein was estimated to be 75kD (lane 2), 110 kD (lane 3), 90 kD (lane 4) and 135 kD (lane 5).

Figure 6: Restriction map of the plasmid pUC924.

The size of the vector pUC9 is 2.7 kb. The cloning site 3' of the lacUV5 ß-galactosidase promoter and operator (PO) contains an EcoRI(E), BamHI(B), SalI(S), PstI(P), and HindIII(H) site. The ß-lactamase gene is indicated by AMP.

The 2.6 kb BglII/EcoRI-fragment of pUC635 is inserted between the BamHI and EcoRI sites. The abbreviation of BglII is "Bg".

Figure 7: Restriction map of the plasmid pMF924.

The 2.6 kb BamHI/HindIII-fragment of pUC924 was inserted into the BamHI and HindIII restriction sites of pEA305 which are located 3' of the hybrid trp-lac promoter (tac) and the aminoterminal coding region of c_1 (λ-repressor).

Figure 8: Restriction map of the plasmid pKK378.

The 3.3 kb BamHI/HindIII-fragment of pUC6130 was inserted into the HindIII-site of the vector pKK2400-11 using a 345 bp BamHI/HindIII-fragment of pBR322 as a linker (which is indicated by a heavy black line). Thus the p138 encoding fragment is located 3' of the hybrid trp-lac promoter (tac) and an ATG start codon.

Figure 9: Secondary structures of p138.

Computer plot of Chou-Fasman calculation of the p138 secondary structure. Additionally, the hydrophobic (closed circles) and hydrophilic (open circles) regions are indicated.

Antigenic sites can be expected in hydrophilic regions with a β-turn. This situation is given in the p60o region and at the carboxy-terminus of the protein.

The regions subcloned into the vectors pUC8 and pUR288 are indicated.

Figure 10: Expression products of bacteria transformed with the plasmid pUR carrying PstI fragments of p138.

A. A coomassie brilliant blue stained SDS polyacrylamide slab gel analysis of lysates of IPTG induced bacteria carrying the various plasmids is shown. Fusion proteins with molecular weights between 120 and 150 kd are indicated with a closed circle. Track M molecular weight markers tracks pUR400-pUR540 lysates of bacteria carrying plasmids containing the regions of p138 as shown in Fig. 3.

B. An enzyme-linked immunoassay of proteins transfered from a gel (similar to that shown in panel A) onto nitrocellulose paper (Western blot) is shown. In this assay a pool of high titered antiserum was used and after washing, the bound immunoglobulins were visualized by sequential reaction with peroxidase coupled to antibodies against human IgG and diaminobenzidine. Only fusion proteins from bacteria containing pUR600 and pUR540 show specific reactions. Plasmid pUC535 (as a positive control) contains almost the whole of p138 coding region, however the protein is unstable and is rapidly degraded. pUC8 is the negative control containing the vector plasmid free from EBV derived sequences.

Figure 11: Expression product of bacteria transformed with the pUC subclones carrying PstI-fragments of p138.

An enzyme-linked immunoassay of proteins electrophoretically transfered from a gel onto nitrocellulose paper (Western blot) was carried out. In this assay a pool of high titered antiserum was used and after washing, the bound immunoglobulins were visualized by sequential reaction with peroxidase coupled to antibodies against human IgG and diaminobenzidine. The fusion protein from bacteria containing pUCP600 was stably produced and shows a specific antigenic reaction.

Figure 12: Construction scheme for plasmid pUCARG1140 encoding both antigenic sites found by expression as ß-gal fusion proteins

a) The 5'-PstI site in pUC600 was removed by digesting with SstI (20bp upstream) and HindIII followed by the ligation with pUC12-SstI/HindIII. From this plasmid the insert was removed with EcoRI and PstI and ligated into pUC8-EcoRI. An oligonucleotide coding for five arginines and two stop codons was inserted into the resulting plasmid pUC601 as single-stranded DNA between the 3'-PstI site and HindIII (pUCARG601). In a last step the 540bp PstI fragment encoding the second antigenic determinant from the C-terminus of p138 was inserted by digestion with PstI and ligation. The resulting plasmid contained both antigenic sites in frame followed by five arginine-residues. It was designated as pUCARG1140.

b) Nucleotide sequence of the oligoarginine linker. The lower strand was synthesized and inserted as a single-strand DNA via bridge formation between the sticky ends of PstI and HindIII.

Figure 13: IPTG-induced expression of the plasmids pUC600, pUC601, pUCARG601 and pUCARG1140 with pUC8 as a control

The upper part shows a Coomassie-stained SDS-PAGE. The newly detected proteins are marked by a black dot. The lower part shows the corresponding western blot obtained after immunostaining with serum from NPC patients. In comparison to pUCP600 the EBV-related protein encoded by pUC601 is about 1.5kD smaller due to the lack of 14 aminoacids (6 amino-acids encoded by the pUC-polylinker and 8 from the PstI-SstI fragment). The size of the protein encoded by pUCARG601 is further reduced for about 11kD since the read through into the lacZ region of pUC is inhibited by stop codons present in the inserted oligonucleotide. In pUCARG1140 the size increases to about 42 kD due to the insertion of the 540bp fragment. The protein is stable in bacterial cells.

Figure 14: Distribution and reactivity of the IgG and IgA antibodies of individual NPC-sera against the two epitopes detected in p138.

Lysates of IPTG-induced E.coli cells carrying the indicated plasmids were independently separated on a 12% SDS-PAGE four-times and the proteins were transferred to Nitrocellulose by Western-blotting. Lanes 1: pUR288 as negative control; lanes 2: pUCARG1140 as a positive control; lanes 3: pUR540; lanes 4: pUR600. Two individual NPC-sera (no. 352 and 354) were incubated with the filters and the bound IgG and IgA antibodies were visualized using peroxidase conjugated anti-human IgG and anti-human IgA rabbit antibodies. The different locations of the proteins in the Western blots, especially of pUCARG1140, result from different electrophoresis times of the SDS-PAGEs.
Whereas in NPC-serum no.352 the main reaction of the IgG and IgA antibodies is directed against the P540 epitope from the C-terminus of p138 (see Fig. 9) in serum no.354 the main part of the anti-p138 antibodies recognizes the P600 epitope (see Fig. 9). This indicates that both antigenic sites are necessary for detecting anti-p138 antibodies in sera.

Figure 15: ELISA test using the protein encoded by plasmid pUCARG1140 as antigen.
Row 1 and 3: EBV-negative sera, row 2: NPC pool serum, row 4-13: individual NPC sera. The dilutions tested are indicated at the bottom; left lane: IgG right lane:IgA.

Figure 16: Purification of proteins carrying oligo-arginine groups at their carboxy-terminus.

A. Sequence of the oligonucleotide encoding five argine residues and two stop codons. A HindIII-site at the 5'-end and a PstI-site at the 3'-end were generated for the insertion of the oligonucleotide into pUC8.

B. Purification scheme of insoluble expressed eukaryotic proteins carrying said Arg-linker at their carboxy-terminus.

Fig. 17:

DNA sequence of the leftward reading frame of the Bam L-fragment encoding gp 250/350.

The coding region for the glycoprotein starts at genomic position 92153 and ends at position 89433. The sequence shown is the respective negative strand, beginning with the BamHI site at position 92703. According to the sequence numbering in this figure the gp 350 encoding region is located between position 556 and 3276. A TATAA-box in the region of basepair 520 is marked with ..., the probable poly-adenylation site at position 3290 with +++. The splice donor and splice acceptor sites are indicated by)(--- for donor and ---)(for the acceptor site. A hydrophobic region near the carboxy-terminus of the coding region is marked with ***. Probably this aminoacid sequence serves as an anchor sequence for fixing the protein to the membrane.

Fig. 18:

Restriction map and open reading frames of the Bam L-fragment

A. Restriction map:
The positions of the restriction enzymes Bam HI, EcoRI, HindIII and PstI are indicated relative to the nucleotide positions of the Bam L-fragment.

B. Open reading frames:
The open reading frames of the Bam L-fragment are indicated as boxes and given for both polarities of the respective DNA sequence.

Fig. 19:

Restriction map of the plasmid pUCLP1.9

The size of pUC8 is 2.7 kb. The cloning region 3' of the LacUV5 ß-galactosidase promoter and operator (PO) contains an EcoRI(E), BamHI(B), SalI(S), PstI(P), and HindIII(H) site. The 1.9 kb subfragment of the Bam L-fragment, indicated by an open bar, was inserted into the PstI site. The reading frame has the same orientation as the lacZ-coding part of pUC8 (indicated by a heavy black line).

Fig. 20:

Restriction map of plasmid pURLP1.9

The vector pUR290 has a length of 5.2 kb and consists of the ß-lactamase gene (AMP^R) and the origin of replication of pBR322. The ß-galactosidase gene is indicated by a heavy black line, the respective promoter-operator region by PO. The restriction enzymes are abbreviated as follows: BamHI(B), ClaI(C), EcoRI(E), HindIII(H), PstI(P), and SalI(S).

The 1.9 kb insert of pUCLP1.9 was introduced between the BamHI and the HindIII site.

Fig. 21:

DNA- and amino acid sequence of the fusion protein encoded by plasmid pURLP1.9

bp 4 - 3069: ß-galactosidase
bp 3070 - 3072: pUR290 linker (given in low letters)
bp 3073 - 3088: pUC8 multiple cloning site (BamHI to PstI; given in low letters)
bp 3089 - 4985: PstI fragment of gp 350
bp 4986 - 4994: pUC8 multiple cloning site (PstI to HindIII; given in low letters)
bp 4995 - END: pBR322 sequence.

Fig. 22:

Expression of the ß-gal : gp 350 protein encoded by plasmid pURLP1.9

Lane 1 and 2 show a coomassie blue stained PAGE of an uninduced (lane 1) and an IPTG induced (lane 2) pURLP1.9 containing clone.

Since there are a lot of bands with different molecular weights, it seems that the main part of the protein is incompletely synthesized.

Lane 3 shows a peroxidase-DAB stained Western blot with NPC sera. It is demonstrated that all newly expressed proteins are antigenic, except that band according to the size of 116 kD which corresponds to the ß-galactosidase.

The bacterial background bands are due to the high content of antibacterial-antibodies in the serum used.

Fig. 23:

Purification of the ß-gal : gp 350 fusion protein encoded by plasmid pURLP1.9

A. Coomassie stained gel; B. Western blot, treated with NPC serum
Lane 1: Uninduced culture
Lane 2: IPTG induced culture
Lane 3: Insoluble proteins of the lysed bacteria, disolved in 8M urea
Lane 4: ß-gal : gp 350 protein containing fractions, pooled after Sepharose 2B-Cl chromatography.

Figure 24: Computer-predicted secondary structure of gp350 comprising the relative values of hydrophilicity (dark circles) and hydrophobicity (grey circles). In the scale given only the loop structures can be seen clearly as line turns of 180°.

Figure 25: Expression of gp350-fragments as ß-gal fusion proteins.
The coomassie blue stained expression products encoded by plasmids pURLEP600 and pURLXP390 are shown in the upper part (pUR288 as control). In the lower part the same probes are shown after immunostaining for demonstrating their reactivity with EBV-positive sera.

Figure 26: Expression of the proteins encoded by pUCLEP600 and pUCARG1230 and their reactivity against EBV-positive sera with pUC8 as control; upper part: comassie-stained SDS-PAGE, lower part immunostained westernblot

Figure 27: Restriction map of the region coding for gp 250/350
The dark bar indicates the region coding for gp 250/350. Furthermore the restriction enzymes used for subcloning, the splice sites, and the inserts of the recombinant expression plasmids constructed according to examples 13 and 15-17 are shown.

Figure 28: DNA sequence and corresponding aminoacid sequence of EBV-related proteins.

A. Protein p54
   Nucleotide sequence and derived aminoacid sequence of protein p54 which is identified in in vitro translation as p47 but correlated with immuno-precipitation with monoclonal antibodies
B. Protein p90
C. Protein p143
D. Protein p150

Figure 29: Expression of the ß-gal::p150 fusion proteins

IPTG-induced clones indicated on top were separated after lysis in an 10 % SDS-PAGE and the proteins were stained with Coomassie-blue. As a control pUR288 was applied to show the size of the ß-galactosidase. All clones produce new proteins larger than the control clone and corresponding to the insert size.

Figure 30: Antigenicity of the ß-gal::p150 fusion proteins

The same lysates from clones shown in Fig. 29 were transferred to nitrocellulose and EBV-related antigens were visualized by immuno staining (see supra). The clone encoding the N-terminal part reacts strongly.

Figure 31: Map of the p150 encoding region

The p150 encoding region is shown as dark bar. The restricion sites used for subcloning and the resulting pUR-clones are also indicated.

Best mode of carrying out the invention

Example 1

Identification of an antigen suitable for diagnosis of NPC

In order to obtain the desired DNA sequences coding for EBV-related antigens of diagnostic significance the following strategy was developed:
Immunoprecipitation of Epstein-Barr viral proteins with various sera from normal adults, patients with fresh infectious mononucleosis or nasopharyngeal carcinoma was used to identify antigens, which are of relevance for the diagnosis of immune status and characteristic for a particular disease (Figure 1). These antigens have been localized on the Epstein-Barr virus genome by hybrid-selected translation. With the use of sequence data, these genes were subcloned from EBV-DNA and expressed in eucaryotic and procaryotic cells.
It was shown by immunoprecipitation that EA and VCA are not single antigens but families of antigens that consist of several polypeptides (G.J. Bayliss, H. Wolf, "The regulated expression of Epstein-Barr virus. III. Proteins specified by EBV during the lytic cycle", J. Gen. Virol 56, p. 105 (1981)).

For the immunoprecipitation the EBV-producing, MA-positive cell line P3HR1, the EBV-positive, non-producing Raji cell line and the EBV-negative cell line BJAB were used. When the cells reached a density of about 10^6/ml, they were diluted with an equal volume of fresh medium. For induction of EBV antigens, P3HR1 cultures were treated with 40 ng/ml phorbol-12-mystrate-13-acetate (modified from zur Hausen et al (H. zur Hausen, F.J. O'Neill, U.K. Freese, E. Hecker, "Persisting oncogenic herpes virus induced by the tumor promotor TPA", Nature 272, p. 373 (1978)) and 3 mM butyric acid immediately after subculture. For the labelling of the proteins, the cells were collected by low-speed centrifugation and resuspended at a density of 2 x 10^6 cells/ml

in methionine-free MEM culture medium containing between 5o and 1oo µCi/ml ^{35}S-methionine. The cells were incubated at 37°C/5 % CO_2 for 4 h and subsequently washed with cold Hanks' phosphate buffered saline (PBS) and resuspended in cold IP buffer (1% Triton-X-1oo, o.1% SDS; o.137 M NaCl; 1 mM $CaCl_2$; 1 mM $MgCl_2$; 1o% glycerol; 2o mM Tris-HCl pH 9.o; o,o1 % NaN_3; 1µg/ml phenylmethylsulphonyl fluoride) at a concentration of 5 x $1o^6$ cell/ml. Then the cells were disrupted by sonication and incubated on ice for 6o min. The extracts clarified by centrifugation at 1oo,ooo x g for 3o min at 4°C.

^{35}S-methionine labelled extracts were immuno-precipitated exactly as described (G.J. Bayliss, et al., supra). The results are shown in Fig. 1.

Antibodies to p138, p1o5, p90 and p80 are present in each of the NPC sera and only in some of the other EBV-infection specific sera. In analogy antibodies to p54 (identical to p58 in G.J. Bayliss et al., supra) are significant for fresh EBV-infection (infectious mononucleosis) as compared to convalescent state. Antibodies to p150, p143, p110 are also present in convalescent sera of healthy individuals and can serve as markers for immunity or, in connection with IgM specific tests for fresh EBV-infection or, in connection with IgA specific tests, for EBV-related neoplasia (NPC and BL).

The next step was to localize the antigens on the EBV genome. Therefore RNA was prepared by lysing the EBV-producing cells described above with 4 M guanidine isothiocyanate and o.5 M 2-mercaptoethanol two days after induction (J.M. Chirgwin, A.E. Przybyla, R.J. MacDonald, W.J. Rutter, "Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease", Biochemistry 18, p. 5294, (1979)). The lysate was centrifuged for one

hour at 20.000 rpm (SW 41, Beckmann) and the supernatant layered on top of 2 ml CsCl density 1.8 g/cm^3. After centrifugation for 17 hours at 150.000 g, the RNA pellet was extracted with chloroform and precipitated with ethanol. 100 µg total cellular RNA was hybridized for 2.5 hours at 52°C in 65 % formamide and 0.4 M NaCl to 16 µg cloned EBV-DNA, which was sonicated, denatured and spotted on small nitrocellulose filters. Bound mRNA was eluted by boiling the filters 90 sec in water. The RNA was translated in vitro with a mRNA dependent rabbit reticulocyte lysate. The translation products were immunoprecipitated using 5 µl of a pool of human NPC sera for one assay after preincubation with a protein extract from unlabelled EBV-negative BJA-B cells as previously described (G.J. Bayliss, G. Deby, H. Wolf, "An immunoprecipitation blocking assay for the analysis of EBV induced antigens", J. Virol. Methods 7, p. 229 (1983)). The immune complexes were bound on protein A-sepharose, washed, eluted by boiling the beads in electrophoresis sample buffer and loaded onto SDS-polyarcrylamid gels. This procedure allowed mapping of a number of viral proteins (Fig. 2) relative to the EBV B95-8 genome. The localization of p138, is given in Fig. 2. Using sequence data (R. Baer, A.T. Bankier, M.D. Biggin, P.L.Deininger, P.J. Fawell, T. J. Gibson, G. Hatfull, G.S. Hudson, S.C. Satchwell, C. Seguin, P.S. Tuffuel, B. Barrell, "DNA-sequence and expression of the B95-8 Epstein-Barr virus genome", Nature 31o, p. 2o7 (1984)), appropriate open reading frames for p138 and p54 were identified (Fig. 2). These open reading frames are completely contained in the right part of the BamA-fragment at the right end of the viral genome.

Example 4

Purification of the ß-gal : p138 fusion protein encoded by the plasmid pUC635

The clone E.coli K12 JM109 pUC 635 was grown at 37°C in 5oo ml L-broth supplemented with Ampicillin as described above until the OD_{560} was o.8. The fusion protein synthesis was induced by IPTG (1 mM) and the incubation was continued for another 2 h. Then the cells were collected by centrifugation for 1o min in a GSA rotor (Sorvall) at 5.ooo rpm and they were resuspended in 5o ml 2o mM Tris-HCl, pH 7.5. For lysating the cells, EDTA (5o mM final concentration) and lysozyme (2 mg/ml final concentration) were added and this mixture was incubated for 3o min at 37°C. In the following, the cells were sonicated (Labsonic 151o, Braun) twice for 8 min, Triton X-1oo was added to a final concentration of 3 % and, after further incubation at 37°C for 3o min, insoluble particles of the suspension were pelleted by centrifugation (SS 34 rotor (Sorvall), 2o min, 1o.ooo rpm). The resulting pellet was dissolved in 2o ml of an 8 M urea, 1o mM Tris-HCl, o,5 % ß-mercaptoethanol, pH 7.5, solution and recentrifuged as before.

Finally 80 mg of the proteins were subjected to a column chromatography (Sepharose 2B-Cl (Pharmacia), length 80 cm, diameter 3 cm) with 8 M urea, 10 mM Tris, 0.1 % ß-mercaptoethanol, pH 7.5, buffer. 30 µl of each of the collected 4 ml samples were analyzed in a 15 % PAGE and the fusion protein-containing fractions were pooled.

Example 5

Cloning of p138 subregions coding for antigenic determinants identified by computer analysis

In principle in diagnostic tests only the antigenic determinant subregions of the antigenic protein are needed. Therefore the p138 amino acid sequence was analyzed by a computer programm and the identified subregions of this gene were introduced in suitable vectors. The production of such small proteins has the advantage that these are less vulnerable to rapid changes of antigenicity with decreasing length of the product. Furthermore especially in conjunction with assays for class specific antibodies they will be of diagnostic value.

According to the method of P. Chou and G. Fasman ("Conformational parameters for aminoacids in α-helical ß-sheet and random coil regions calculated from proteins", Biochemistry 13, p. 211 (1974)) the calculation of the appropriate secondary structure of a protein caused by its aminoacid sequence (primary structure) is possible. Superimposed on the suggested structure, the program determines the relative hydrophilicity and hydrophobicity. Both data sets are combined and a computer graphic is drawn that shows α-helical, ß-sheet, ß-turning and randomly coiled regions of the secondary structure. Thereby the hydrophilic and hydrophobic regions are shown as open and closed circles, respectively.

An example of such computer graphic is shown for the p138 amino acid sequence in Fig. 9.

All expressed proteins display almost the expected size, but the yield varied over a wide range. The proteins encoded by pUC635 and pMF924, seem to be more stably expressable than the non-fusion proteins from pUC924 and pKK378. However, the amount of even the highest expressed protein from pUC635 is too low for a large-scale production since in the Coomassie-stained gel only a very weak band was visible which may be due to the large size of the eukaryotic protein.

Example 3

Immunological assay of the proteins encoded by pUC635 pUC924, pMF924 and pKK378

The host cells transformed with plasmids pUC635, pUC924, pMF924 and pUK378 were cultivated in L-broth supplemented with 50 µg/ml Ampicillin to a cell density of D_{600}=0.8. Then, for the induction of the ß-galactosidase the lactose analogon isopropyl-ß-D-thiogalactopyranoside (IPTG; Sigma) was added (final concentration: 1mM). After a further incubation of 1.5 h at 37°C, 1.5 ml of the culture were centrifuged. The bacteria were resuspended in 2oo µl boiling mix (2 % SDS, 5 % mercaptoethanol, 3 % sucrose, 5o mM Tris-HCl, pH 7.o) and heated for 1o min at 1oo°C. 2o µl of the resulting protein extract were separated on a 12.5 % polyacrylamide gel and finally the proteins were visualized by coomassie-blue staining, but since the yield of the expression product is very low, an immunostaining was necessary. Therefore the electrophoretically separated proteins were transferred to a nitrocellulose filter, i.e. a "Western-blot" was prepared (J. Renart, J. Reiser, G.R. Shark "Transfer of proteins from gels to diazobenzyl-methyl-paper and detection with antisera",

Proc. Natl. Acad. Sci. USA 76, p. 3116 (1979), S. Modrow, H. Wolf, "Characterization of herpesvirus saimiri and herpesvirus ateles induced proteins", in: Latent Herpes Infections in Veterinary Medicine, Martinus Nijnoff Publ., p 105 (1984)).

The Western-blot was prepared with a current intensity of 0.8 A for 3 h in Western-blot buffer (72 g glycin, 15 g Tris, 1 l methanol, H_2O dest. ad 5 l). Then the nitrocellulose was saturated with Cohen buffer for 3 h (0.1 % Ficoll 400, 1 % polyvinylpyrrolidone, 1.6 % BSA, 0.1 % NP40, 0.05 % gelatine, 0.17 M H_3BO_3, 28 mM NaOH, 150 mM NaCl, 6 mM NaN_3, pH 8.2) and incubated overnight with 1:50 diluted high titered EBV specific serum from NPC-patients. The serum had been preabsorbed to a bacterial-protein extract (1 ml/10^9 E. coli cells) to reduce the bacterial protein generated background. Afterwards unbound IgG was removed by washing the nitrocellulose filter for 5 h in gelatine buffer (50 mM Tris-HCl, 5 mM EDTA, 150 mM NaCl, 0.25 % gelatine, 0.5 % Triton, 0.2 % SDS, pH 7.5). For visualizing the blotted EBV-specific proteins rabbit anti-human-IgG-antibodies coupled to peroxidase and diluted 1:200 in TN buffer (154 mM NaCl, 10 mM Tris, pH 7.4) was added. After 2 h at RT, unbound rabbit antibodies were removed by washing with gelatine buffer as described above. Finally the peroxidase reaction was carried out in 100 ml 50 mM Tris-HCl, pH 7.5, by adding 50 mg diamino-benzidine (Sigma) and 40 µl H_2O_2 and incubating 10 min. at RT. The results of this experiment are shown in Fig. 5.

(5 g yeast extract, 1o g tryptone, 5 g NaCl) incubated 1.5 h at 37°C, and finally plated on L-broth agar-plates (1.5 %) supplemented with 5o µg/ml Ampicillin (Sigma) and 4o µg/ml X-gal (Boehringer). During this incubation bacteria carrying religated pUC8 molecules yield blue colonies and those which carry recombinant plasmids yield white colonies.

For identification of clones that carry the desired recombinant plasmid, twelve white colonies were picked and grown overnight at 37°C in L-broth. Aliquots of DNA-preparations according to H.C. Birnboim and J. Doly ("A rapid alkaline extraction procedure for screening recombinant plasmid DNA", Nucl. Acids Res. 7, p. 1513 (1979)) were digested by BamHI and HindIII and electrophoresed on an agarose gel as described before. Furthermore, for demonstrating the orientation of the integrated fragment, a digest was carried out with BamHI and BglII. Finally the 3.3 kb was checked by a XhoI digest.

Plasmid pUC635 carries the 3.0 kb XhoI-subfragment of the BamA-fragment (pBR322 BamA) in the proper orientation and the proper reading frame relative to the lac UV5 promoter and is used for the expression of nearly the whole p 138 (Fig. 4). The fusion protein encoded by pUC635 is composed of 12 amino acids of the ß-galactosidase amino terminus, about 1o2o amino acids of p138, 6o amino acids of the carboxy terminal part of the ß-galactosidase and another 29 amino acids of a pBR322 encoded region. Plasmid pUC6130 carries the 3.3 kb fragment in the opposite orientation (Fig. 4) Since the strain E.coli K12 JM83 is not a ß-galactosidase repressor overproducer, the fusion protein is constitutively expressed. Therefore the plasmid pUC635 was introduced into the ß-glactosidase repressor overproducer strain E.coli K12 BMH71-18 (DSM 3413) (U. Rüther, B. Müller-Hill, "Easy identification of cDNA clones", EMBO Journal 1o, p. 1791

(1983)). Instead of strain E. coli K12 MBH71-18 strain E.coli K12 JM109 (DSM 3423) can also be used (without essential alteration of the experimental procedure).

Besides pUC635 three other plasmids were constructed:pUC924, pMF924 and pKK378 (Figures 6 to 8).

The insert of pKK378 starts at the same XhoI-site and continues up to the third XhoI-site located 250bp 3' of the stop codon. This fragment of 3.3kb was generated by an incomplete digest and inserted behind the tac-promotor and the start codon of pKK240-11 (F.Amann et al., supra). The expression product contains only two bacterial amino acids and its size is smaller then the size of the expression product of pUC635 because the bacterial lacZ part is missing.

pUC924 contains the fragment from the Bgl II-site to the third Xho I site. pUC9 (DSM 3421) was used as vector. Since the size of the insert is smaller than in pUC635 and since the stop codon from p138 is used, the molecular weight of the expression product is expected to be smaller than in pUC635 and pKK378.

The plasmid pMF924 was constructed from pEA305 (E. Amann et al., supra) and the same BglII-XhoI fragment as in pUC924. pEA305 has a tac-promotor followed by the N-terminal part of the C1repressor, the resulting fusion protein is expected to be 17kd larger than in pUC924.

These constructs were tested for the production of EBV-related antigens by inducing the tac- and lac-promotors with IPTG and separating the proteins on an SDS-PAGE. None or only weak new bands could be detected on Coomassie-blue stained gels in the regions with the expected sizes. But after a transfer of the proteins onto nitrocellulose and immunostaining with a high titered NPC-pool serum and a peroxidase conjugated second anti-IgG antibody new EBV-specific bands were clearly detectable in all constructs.(Fig. 4)

Example 2

Cloning of the p138 encoding region

According to the sequence data of R. Baer et al., (supra), there is a large open reading frame contained in the BamA-fragment of EBV B95-8 which is suitable for encoding p138. The nucleotide sequence, the corresponding aminoacid sequence and the respective regulatory elements of the gene of p138 are given in Fig. 3.

50 µg DNA of the plasmid pBR322-BamA (J.Skare et al., supra) were digested with 50 U XhoI (Boehringer) for 2 h at 37°C in a total volume of 150 µl containing 150 mM NaCl , 10 mM $MgCl_2$, 6 mM mercaptoethanol, 6 mM Tris-HCl, pH 7.9. 30 µl stop buffer (10 mM Tris-HCl, 50 mM EDTA, 60 % sucrose, 1 % bromphenolblue, pH 7.5) were added, the mixture was put onto a preparative 1 % agarose gel in acetate-buffer (0.04 M Tris-acetate, 2 mM EDTA, pH 7.6), and electrophoresed for 16 h with 40 V at 4°C. As a size marker HindIII digested λ-phage DNA (Boehringer) was used. After staining the gel in Tris-acetate buffer with ethidium bromide (0.5 µg/ml) for 1 h at room temperature (RT), the DNA was visualized by UV-illumination and the bands corresponding to 3.0 and 3.3 kb were excised (the 3.0 kb XhoI-generated fragment is the desired fragment, the 3.3 kb XhoI-generated fragment is a partial digest product (one XhoI restriction site was not cut)).

The DNA of the bands was eluted by putting the agarose pieces into dialysis bags, adding 3 volumes of Tris-acetate buffer and electrophoresed for 4 h (100 V, 4°C). Further purification was carried out by a chromatography with Elutip D columns (Schleicher & Schuell) according to the procedure recommended by the manufacturer, extraction of the contained ethidium bromide with isoamylalcohol and pre-

cipitation of the DNA by adding 2.5 volumes ethanol and incubating overnight at -2o°C. The DNA was collected by centrifugation in a Sorvall SS 34 rotor (17.ooo rpm, 2o min) and washed with 7o % ethanol. After lyophilization the DNA was dissolved in 15 µl TE buffer (1o mM Tris-HCl, 1 mM EDTA, pH 7.5).

The DNA concentration of the two isolated fragments was estimated by electrophoresing 1 µl each in parallel with 1oo ng and 1 µg of pUC8 DNA.

SalI digested DNA of the vector pUC8 (deposited with the Deutsche Sammlung für Mikroorganismen (DSM), Göttingen , West Germany, under the accession number DSM 3420) (J. Messing, J. Vieira, of double-digest restriction fragments", Gene 19, p. 269 (1982)) was prepared as described before, except that for inhibition of religation of the vector during the following ligase reaction the DNA was treated with alkaline phosphatase (o.5 units (Boehringer), 3o min at 37°C).

In the following, the two purified fragments were each inserted into the cleaved vector (SalI and XhoI produce the same cohesive ends, i.e. -TGCA-). For this purpose for each of the fragments a ligation reaction was carried out with 3oo ng fragment DNA and 1oo ng pUC8 DNA in a total volume of 2o µl ligase buffer (1o mM Tris, 1o mM $MgCl_2$, 6 mM mercaptoethanol, o.6 mM ATP, pH 7.5) containing 1U T4-DNA ligase (Boehringer). After 2o h at 14°C, 3o µl TE buffer and 2oo µl competent E.coli JM83 cells (ATCC 35607) (J. Vieira, J. Messing, "The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers", Gene 19, p. 259 (1982)) were added. The transformation was done according to the calcium chloride procedure (M. Mandel, A. Higa, "Calcium dependent bacteriophage DNA infection", J. Mol. Biol. 53, p. 154 (197o)). Then the cells were mixed with 1.5 ml L-broth

Based on the assumption that antigenic sites are mainly located in hydrophilic ß-turns which are located on the surface of the protein, the region between about amino acid 520 and the carboxy-terminus of p138 should be antigenic. The corresponding DNA sequence is represented by a PstI-fragment of pUC635.

Thus pUC635 was cleaved with PstI and all PstI-fragments were isolated and introduced into PstI-cleaved pUC8, the remaining vector fragment with additional 400 bp (up to the first PstI-site of the p138 coding sequence) was religated (all methods as described in example 2).

The resulting recombinant plasmids were designated pUC P400, pUC P380, pUC P600, pUC P210, pUC P750, and pUC P540, respectively.

The aminoterminal region of the p138 encoding sequence was cloned by digesting the plasmid pBR322-BamA with PstI and HgiAI and inserting said fragment into PstI cleaved pUC9. (J. Messing et al., supra) (methods as described in example 2). The resulting recombinant plasmid is designated pUC HP.

With the exception of pUC HP in which translation stops at the 3' end of the insertion, in all subclones orientation and reading frames relative to pUC8 are correct.

Finally the recombinant plasmids were introduced into E. coli K12 JM 109 cells.

Example 6

Expression of antigenic determinants identified by computer analysis using pUR288 p138 subclones

Since pUC subclones (example 5) sometimes are not stably expressible in bacteria, because they cannot build up a suitable tertiary structure due to their shortness and therefore can be degraded by proteases to a larger extent than complete proteins we constructed recombinant plasmids encoding large fusion proteins using at least a part of the β-galactosidase encoded by pUR288 (DSM 3415) (U. Rüther et al., supra) by cleaving said PstI-fragment subclones with BamHI and HindIII, isolating the respective fragments and ligating them into BamHI and HindIII cleaved pUR288 (all methods as described in example 2).

The expression was carried out in E.coli K12 JM109. The products were analyzed as described in example 3. After coomassie-blue staining of the gel several large fusion proteins of different size were detected, however, after preparation of a Western-blot, only the products expressed by pUR600 and pUR540 showed specific reaction with the IgG antibodies mentioned (Fig. 10).

These results are in good agreement with the computer analysis.

Additionally the expression of the clones obtained according to example 5 was carried out according to example 3. The products, too, were analyzed as described in example 3. From the coomassie-blue stained gel it can be taken that only plasmids pUCP600 and pUCP380 code for a stable fusion protein. The Western-blot shows that only pUCP600 derived fusion protein is antigenic (Fig. 11). This fusion protein contains 11 amino acids encoded

by the aminoterminal cloning site, a region encoded by about 600 bp of p138 and carboxyterminal amino acids of the lacZ gene. Thus, the recombinant expression plasmids pUR600 and pUR540 as well as pUCP600 can be used for the production of large and small fusion proteins, respectively, containing an antigenic determinant of EBV-protein p138.

Example 7

Application of the protein encoded by plasmid pUCP600 for the stabilization of per se unstable parts of eukaryotic proteins

By means of the experiments of example 6 it was shown that the p138-derived protein parts (regions) are unstable with the exception of the protein encoded by plasmid pUCP600. The second antigenic region from the C-terminus of p138 (P540, see Fig. 9) is not stably expressible using the recombinant pUC-vector pUCP540.
The ability of the P600-region of p138 to stabilize such a per se unstable expression product is shown in this example.

For this purpose it was necessary to remove the 5'-PstI-site of pUCP600 by digesting the plasmid with SstI and HindIII(the SstI site is located about 20 bp 3' from the first PstI site). The p138-related SstI-HindIII fragment was inserted into SstI/HindIII cleaved pUC12 (DSM3422) J. Messing, "New M13 vectors for cloning", in Methods of Enzymology Vol. 101, Part C., R. Wu, L. Großmann and K. Moldoave (eds.), Acad. Press, New York, 1983, 20-78).

Then the resulting recombinant plasmid was digested with EcoRI and PstI. The obtained 600 bp fragment was inserted into plasmid pUC8. The 5'-PstI site was now replaced by an SstI site and thus the reading frame is reconstituted at the 3'- and the 5'-end of the insert (Fig. 12a). The resulting recombinant plasmid pUC601 still expresses a stable product (Fig. 13).

Between the PstI and HindIII site at the 3'-terminus of the EBV-encoded sequence a synthetic oligonucleotide obtained according to known methods coding in frame for 5 arginine and 2 stop codons was inserted as shown in Fig. 12b). The resulting plasmid pUCARG601 encodes the P600 region of p138 fused at its C-terminus to 5 arginine residues.

In a last step the PstI fragment encoding the P540 region of p138 was ligated to the PstI fragment encoding the P600 region of p138 after digestion with PstI. The resulting recombinant plasmid pUCARG1140 encodes a stable protein of about 43kd which contains two antigenic sites of p138 fused in frame. In this fusion protein the protein region P600 stabilizes the protein region P540 (Fig. 13). The arginine residues at the ~~carboxy~~terminus of the expression product may be used for the purification of the resulting fusion protein as described by Sassenfeld and Brewer (supra) (Fig. 16).

Example 8

Construction of the recombinant plasmid pUCARG680

From the plasmid pUCARG1140 a modified version was constructed which lacks 435bp of the p138 encoding region, the C-terminal part of the p600 fragment and the N-terminal part of the p540 fragment. The main antigenic

sites predicted by the computer program are still present. The plasmid was designated as pUCARG680 and its construction was achieved by digesting pUCARG1140 with NcoI (cleavage site coresp onds to bp1841 and bp3243 in Fig. 3). Since the reading frames in the p600 NcoI site and the p540 NcoI site do not fit, the sticky ends were removed with S1-Nuclease.

30 µg of pUCARG1140 were digested with NcoI, the 3.3kb vector-p138 fragment was separated by gelelecrophoresis and purified. 5µg of this DNA fragment were digested with 100 units S1-Nuclease for 15 min at roomtemperature in 100 µl containing 33mM Na-acetate, 50 mM NaCl, 0.03mM $ZnSO_4$, pH 4.5. The digest was stopped by phenol extraction. After precipitation with ethanol the DNA was religated with T4-DNA ligase and used to transform competent E.coli K12 JM109 cells. The resulting clones were screended for the appearance of a new protein with 30kb in size (pUCARG680). The shortened p600/p540 fusion protein encoded by pUCARG680 still reacts as an antigen.

The newly constructed recombinant plasmid pUCARG680 was deposited with the DSM under the deposition number DSM3408.

Example 9

Assay of the antigenicity of the fusion protein encoded by plasmid pUCARG1140

Immunoblots with the fusion proteins encoded by the recombinant plasmids pUCARG1140, pUR540, and pUR600 (examples 6 and 7) using individual NPC-sera reveal that the immunological reactions differ in various patients (Fig. 14). In this context it has to be understood that said plasmids encode fusion proteins containing the p138 regions P540 + P600, P540, and P600, respectively (see Fig. 9).

Whereas in NPC serum no.352 the main fraction of the IgG and IgA antibodies is directed to the P540 region, the main fraction in NPC serum no.354 is directed to the P600 region of p138. A representative pool prepared from many sera from NPC patients did not detect additional antigenic sites. The conclusion from this finding is that the antigenic determinants P540 and P600 as encoded by the recombinant plasmids of the present invention are necessary and sufficient to achieve the desired specificity for ELISA tests useful for diagnostic purposes.

Example 10

Application of plasmid pUCARG1140 encoded fusion protein for the detection of NPC in ELISA tests

The purified fusion protein encoded by pUCARG1140 was coated on micro-titer plates. Ten individual NPC-sera were tested for their IgG and especially for their IgA reactivity. The IgA-anti-EA titer of these sera was previously determined in conventional immunofluorescence tests. The highest titer found was 1:80. In the ELISA test shown in Fig. 15, two EBV-negative, one NPC-serum pool and ten individual NPC-sera were tested up to a dilution 1:10640. The test was performed according to the usual ELISA protocol. Bound antibodies were detected with peroxidase conjugated mouse anti human IgG, i.e. IgA and peroxidase reaction. All NPC sera show a reaction with the coated antigen (up to 1:2560 in IgA) and no background reaction could be observed in the negative controls. This result indicates that the pUCARG1140 encoded expression product is suitable for the diagnosis and early detection of NPC.

Example 1T

Cloning of a subregion of the gene coding for gp 350 in the vector pUC8

The coding region of gp 250 and gp 350 was mapped to the Bam L-fragment (J. Skare et al., supra) of the EBV B95-8 genome. As both polypeptides share identical regions it was supposed that both proteins are encoded by overlapping reading frames (M. Hummel, D. Thorley-Lawson, E. Kieff, "Epstein-Barr virus DNA fragment encodes messages for the two major envelope glycoproteins (gp 350/300 and gp 220/200)", J. of Virol. 49, p. 413 (1984)). The sequence data of Baer et al. (supra) revealed a large open reading frame including a donor splice site and an acceptor splice site in said Bam L-fragment of the virus genome (Fig.17 and 18).

It is assumed that gp 350 is the translation product of the unspliced mRNA transcribed from this region and gp 250 is a product of the corresponding spliced mRNA (Fig.17). Since both products are found in the viral capsids it is assumed that a differential splicing of said mRNA in a manner comparable with the immunoglobulin heavy chain genes (T. Honjo, "Immunoglobulin genes", Ann. Rev. of Immunol. 1, p. 499 (1983)) takes place. During this splicing 630 bp of the mRNA coding for gp 350 are removed to yield the gp 250 coding mRNA (Fig. 17 and 27 (dotted lines)) (R. Baer et al., supra).

Therefore the whole or a part of the reading frame of gp 350 was cloned for finally isolating and producing a gp 350 related product. It should be kept in mind, that not only gp 250 but also gp 350 are highly glycosylated proteins. In contrast, the proteins produced by expression of the recombinant DNA molecules according to the present

invention differ from the respective viral proteins normally occuring in nature. If expression is carried out in prokaryotes unmodified proteins are obtained whereas expression in eukaryotes gives proteins with different patterns of glycosylation or else modifications as compared to the natural product.

The Bam L-fragment was introduced in pBR322, and E. coli K 12 HB 101 was transformed with the recombinant plasmid obtained. (J. Škare, et al., supra)
Instead of the host E.coli K12 HB101 the host bacteria used in the present invention can also be used.

The contents of the publications of M. Hummel et al. (supra), J. R. North et al. (J. R. North, A. J. Morgan, J. L. Thompson, M. A. Epstein, "Purified Epstein-Barr virus M_r 340.000 glycoprotein induces potent virus-neutralizing antibodies when incorporated in liposomes", Proc. Natl. Acad. Sci. USA 79, p. 7504 (1982)) and D. A. Thorley-Lawson and C. A. Poodry ("Identification and Isolation of the Main Component (gp350-gp220) of Epstein-Barr Virus Responsible for Generating Neutralizing Antibodies In Vivo", J. Virol. 43, p. 730 (1984) do not permit predictions that subregions of the gp 250/350 encoding sequence are coding for sufficiently antigenic and/or immunogenic proteins and that these products after selective introduction of these subregions can be stably expressed in prokaryotic and eukaryotic cells. It is therefore surprising that completely unmodified or in a different way modified gp 250/350 related proteins of the present invention are sufficiently active antigens and/or immunogens. In particular in previous publications it was not excluded that minor carbohydrate residues of the protein contribute significantly to the antigenic or immunogenic potential of this protein.

As shown in Fig.17, a 1.9 kb PstI-PstI-fragment of the Bam L-fragment (Hummel et al., supra) contains the part of the gp 350 coding region beginning at aminoacid position 232 and ending at aminoacid position 825.

A large scale preparation of the pBR 322-BamL plasmid DNA was done according to the method published by H.C. Birnboim and J. Doly ("A rapid alkaline extraction procedure for screening recombinant plasmid DNA", Nucl. Acids Res. 7, p. 1513 (1979)). 50µg of this DNA were digested for 2 hours at 37°C with 100 units PstI (Boehringer) in 50 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 10 mM Tris-HCl, pH 7.5. The digestion was stopped by addition of 1/5 vol. 50 mM EDTA, 60 % sucrose, 2 % bromphenolblue. The resulting solution was electrophoretically separated on a 1 % agarose gel (Seakem, FMC) in Tris-acetate buffer (0.04 M Tris-acetate, 2 mM EDTA, pH 7.6). As a size marker HindIII digested λ-phage DNA (Boehringer) was used. After the electrophoresis at 40 V for 14 hours at room temperature. (RT), the gel was stained in Tris-acetate buffer containing 0.5 µg/ml ethidium bromide.

The DNA bands in the gel were visualized by UV-illumination and the 1.9 kb PstI-PstI fragment was isolated as described in example 2.

PstI digested DNA of the vector pUC8 was prepared as described before, except that for inhibition of religation of the vector during the following ligase reaction, the DNA was treated with alkaline phosphatase (0.5 units (Boehringer), 30 min at 37°C).

The concentration of the purified fragments was estimated by electrophoresing 1 µl each in parallel with 100 ng and 500 ng of pUC8-DNA (under conditions described above).

400 ng of the 1.9 kb PstI-PstI-fragment and 100 ng of the PstI digested vector DNA were ligated, E.coli K12 JM109 was transformed with the ligated plasmid DNA and positive clones were identified as described in example 2.

The obtained clone was designated E. coli K12 JM109 pUCLP1.9 and the resulting recombinant plasmid pUCLP1.9, respectively.

Example 12

Cloning of a subregion of the gene coding for gp 350 in the vector pUR290

For the expression of a stable product of the gp 350 subregion said 1.9 kb PstI-PstI-fragment was reclon in the vector pUR290 (DSM 3417) (Fig. 20) (U.Rüther et al., infra). The resulting recombinant plasmid is coding for a fusion protein of an aminoterminal region of the ß-galactosidase, followed by the aminoacids 232 to 825 of gp 350 and aminoacids coded by the cloning-site of pUR290 and pBR322 nucleotide residues. The respective aminoacid sequence is given in figure 21.

50 µg DNA of the plasmid pUCLP 1.9 were digested with 100 units BamHI and HindIII and separated on a 1 % agarose gel as described above. The resulting 1.9 kb BamHI/HindIII fragment that contains only a few more nucleotides than the PstI-PstI-fragment originally introduced into pUC8 was separated from the other resulting fragments on a 1 % agarose gel (as described above). Finally it was isolated from the gel as described above and ligated into BamHI/HindIII digested DNA of the vector pUR290 (U. Rüther, B. Müller-Hill "Easy identification of cDNA clones", EMBO Journal 1o, p. 1791 (1983)) according to the methods described above.

The next step was the transformation of the ß-galactosidase repressor-protein overproducer strain E. coli K12 JM109 with these recombinant DNA molecules. The transformants were plated and analysed as described above, except that the aliquots of the DNA preparations were digested with BamHI/HindIII and EcoRI. The resulting clone, E. coli K12 JM109 pURLP1.9 carries the plasmid pURLP1.9, that is a recombinant of said BamHI-HindIII 1.9 kb fragment of the plasmid pUCLP1.9 and the vector pUR290 (see Fig. 20).

Example 13

gp 350 related polypeptides synthesized by E. coli K12 JM109 pURLP1.9

In an overnight culture E. coli K12 JM109 pURLP1.9 was grown at 37°C in 5ml L-broth supplemented with 50 µg/ml Ampicillin. The culture was then diluted to an optical density at 560 nm(OD_{560}) of 0.4, and 4 ml of this bacteria suspension were incubated at 37°C until an OD_{560} of 0.8. The expression of the genetic information carried by plasmid pURLP1.9 was then induced as described in example 3 and finally the proteins were visualized by coomassie-blue staining as described in example 3.

In comparison with the control experiment, several new proteins, encoded by the plasmid pURLP1.9 and ranging in size from 116 kD to 200 kD, were detected (Fig. 22). The different size of the expression products may be due to incomplete mRNA synthesis or translation. To prove that the new proteins are EBV-related products, all the electrophoretically separated proteins were transferred to a nitrocellulose filter, i.e. a "Western-blot" was prepared according to the method described in example 3. The results of this experiment are shown in Fig. 22.

Example 14

Purification of the ß-gal : gp 350 fusion protein encoded by the plasmid pURLP1.9

The replacement of the natural carboxyterminal amino acid sequence of the ß-galactosidase by a gp 350 related amino acid sequence prevents the formation of ß-galactosidase tetramers. Furthermore the newly expressed fusion protein is present in a high concentration in the bacterial cell. Therefore the fusion protein precipitates in the cytoplasm of the host cell.

According to the method described in example 4 the clone E. coli K12 JM109 pURLP1.9 was used for the production of the corresponding fusion protein.

The results of the several stages of this purification procedure are shown in Fig. 23.

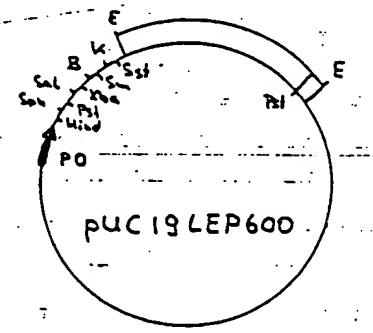
Example 15

Expression of selected antigenic epitopes of gp250/350 as ß-galactosidase fusionproteins

Fig. 24 shows the computer-predicted secondary structure of gp350 together with the relative values of hydrophilic (dark circles) and hydrophobic (grey circles) areas. ß-turns or loop structures are indicated as line turns of 180° (α-helices, ß-sheet and coil structures are barely discernable in the scale used). Based on the assumption that antigenic sites are mainly located in ß-turns in an hydrophilic environment, which may be exposes to the surface of the protein, the regions at about aminoacid 50 and aminoacid 740 and 800-830, respectively, are expected to represent antigenic epitopes.

Subcloning and expression of the N-terminus of gp250/350

The EBV BamHI-L fragment which was cloned in pBR322 (see J. Skare et al., supra) was digested with EcoRI (restriction sites at positions 650 and 1284 in the sequence given in Fig. 17), the resulting 634 bp fragment was eluted from an agarose gel after electrophoresis and ligated to EcoRI linearised pUC19 (DSM3425) (Yanisch-Perron et al. / Gene 33, 103-119 (1985)). Then, E.coli K12 JM109 was transformed with the ligation products (all steps were carried out as described in Example 2). According to example 2 the recombinant plasmids obtained were tested for the orientation of their insert using suitable restriction enzymes. A recombinant plasmid carrying the insert in the opposite orientation of the reading frame relative to the reading frame of the lacZ gene of the pUC19 plasmid was designated as pUC19LEP600 and used for further cloning:



pUC19LEP600 was digested with BamHI and PstI (the BamHI site is derived from pUC19, the PstI site corresponds to position 1248 in Fig. 17), the resulting 600bp fragment was inserted into pUR291 (DSM3418) (Rüther, supra), previously digested with BamHI and PstI. The resulting recombinant plasmid pURLEP600 displayed the following sequence in its linker region at the C-terminus of the ß-galactosidase:

pUR291 / pUC19 / gp250/350

ß-gal-TGT CGG GGA TCC CCG GTA CCG GAG CTC GAA TTC CCA TTT---------- ACC

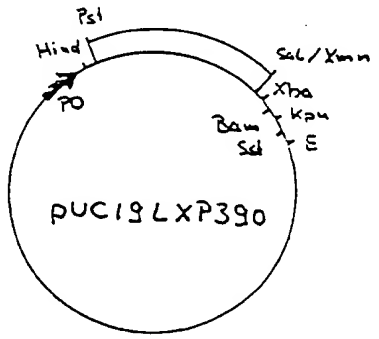
/ pUR291

TGC AGC CAA GCT TAT CGA TGA

The expression of the fusion protein from this recombinant plasmid after IPTG-induction was carried out as described in example 3. The result of this experiment is shown in Fig. 25. From Fig. 25 (lower part) it can be taken that the expression product obtained is recognized as a moderately antigenic protein by a pool of NPC-sera.

Subcloning and expression of the C-terminus of gp250/350

The region covering the antigenic epitopes near the C-terminus which, according to the computer-directed analysis, also is expected to be antigenic, was isolated by digesting the plasmid pUCLP1.9 (see example 11) with XmnI (restriction site at position 2760 in Fig. 17) and HindIII (restriction site in the region derived from the pUC-plasmid). The purified 386 bp fragment was inserted into pUC19 previously digested with HincII and HindIII. The resulting plasmid which was introduced into E.coli K12 JM109 is pUC19LXP390:



The insert of pUC19LXP390 was cut out with BamHI and HindIII and ligated into pUR288 digested with the same enzymes. The resulting recombinant plasmid was introduced into E.coli K12 JM109 and was designated as pURLXP390. The sequence in its linker region is as follows:

```
        pUR288      /      pUC19       /        gp350          /  pUC8  /
ß-gal-TGT  CGG GGA TCC TCT AGA GTC AGT TCC CAC-------GTA CTG CAG CCA AGC
pUR288
TTA TCG
```

After IPTG-induction a ß-galactosidase fusion protein was synthesized by said transformed host. In a Western blot the expression product shows a high reactivity with the NPC sera pool (see Fig. 25, lower part).

Example 16

Use of the ß-gal::gp250/350 fusion proteins encoded by the newly constructed recombinant plasmids in diagnostic tests

Jilg et al. (W.Jilg and H. Wolf, "Diagnostic Significance of Antibodies to the Epstein-Barr Virus-Specific Membrane Antigen gp250", The Journal of Infectious Diseases, 152, 222-225(1985)) have shown the validity of gp250 and gp350 as antigens for the determination of the immune status to EBV and especially for the diagnosis of chronic EBV-infections. Persons showing a normal immune response after an EBV-infection possess antibodies against gp250 and gp350, whereas patients suffering from chronic EBV-infection show an immune response only to gp350 which still contains the additional intron sequence (see Fig. 27). The serological status of these persons can be checked in ELISA tests using the three fusion proteins,

purified according to the method given in example 4. An antibody reaction to all three fusion proteins indicates a normal immune status. If there is nor or weaker reaction with the proteins encoded by pURLEP600 and pURLXP390, but reactivity against pURLP1.9 (which contains the intron sequences, see Fig. 27) a chronic EBV-infection is very likely.

IgA antibodies to the membrane protein gp250/350 and to subfragments thereof are absent in the normal population, but present in 58 % of Nasopharyngeal Carcinoma patients when measured in a relatively insensitive immunofluorescence assay. These results are similar to the detection rate of IgA antibodies to EBV specific early antigens in comparable testsystems. In analogy the more sensitive ELISA test brings the detection rate close to 100 % with only minimal increase of false positive results. Therefore the antigens encoded by the newly constructed recombinant plasmids pURLEP600, pURLXP390, and pURLP1.9, respectively, are valuable substances for the initial diagnosis and the control of a therapy of Nasopharyngeal Carcinoma.

Example 17

Expression of the N-terminal gp250/350 fragment in the plasmid pUC8

The recombinant plasmid covering the N-terminal region of gp250/350, pUC19LEP600 (see example 15), was digested with BamHI and PstI. The EBV derived fragment was isolated and ligated into pUC8, previously digested with the same enzymes. The sequence in the linker region of the resulting clone, pUCLEP600, is the following:

pUC8 / pUC19 / gp350

ATT ACG AAT TCC CGG GGA TCC CCG GGT ACC GAG CTC GAA TTC CCA

pUC8

TTT-------ACC TGC AGC CAA GCT TAT

After induction with IPTG, the fusion protein encoded by pUCLEP600 is quite stable in the bacterial cells and is recognized as an antigen by the NPC sera pool (see Fig. 26). The bacterial fusion part consists of 14 amino-acids at the N-terminus and 9 at the C-terminus. The value of this protein is its applicability in a vaccine, especially when it is fused with the per se instable second antigenic region from the C-terminus as it was determined with the β-gal fusion proteins (see example 15).

The inserts of the recombinant expression plasmids and cloning plasmids constructed according to examples 11 and 15 to 17 are summarized in Fig. 27.

Example 18

Expression of the N-terminal part of gp250/350 as a p138::gp250/350 fusion protein

The plasmid pUC19LEP600 (see example 15) was digested with PstI and the resulting 600bp fragment was ligated to the PstI linearised plasmid pUCARG601 (see example 7). The gp350-insert was checked to be in the same orientation as the pUCARG601-reading frame and the resulting recombinant plasmid was designated as pUCARG1230. The sequence in the linker region and at the junction sites of the obtained plasmids is the following:

1

pUC8 / pUC12 / p138 / pUC19

ATG ACC ATG ATT ACG AAT TCG AGC TCT CTG ACC-------ATC CTG CAG GTC GAC TCT AGA

/ gp350 / pUCARG601

GGA TCC CCG GGT ACC GAG CTC GAA TTC CCA TTT-------ACC TGC AGC GTC GTC GTC GTC GTT GAT AAC GTT

After induction with IPTG, E.coli K12 JM109 carrying pUCARG1230 expresses a stable and antigenic protein which consists of antigenic regions from two different proteins, namely p138 and gp250/350 (see Fig. 26).

Furthermore it can be used as antigen in ELISA tests and also for vaccination.

Example 19

Neutralisation test with sera derived from rabbits immunized with gp250/350 antigens

Supernatants from B95-8 cells were used to immortalize human umbilical cord blood cells (Lymphocyte fraction from Ficol/Hypaque gradient). $0,5 \times 10^6$ lymphocytes were seeded per 0,5 ml microtiter plate well and 50 µl of a cell-free supernatant of B95-8 cells were added and allowed to adsorb for 2 hours at 37°C. After incubation the virus-containing medium was removed, cells were washed with RPMI1640 medium containing 10 % fetal calf serum and incubated in 200 µl of the same medium at 37°C

in a 5 % CO_2 atmosphere. Developing colonies of lymphoblastoid cells were evaluated not sooner than three weeks after the start of the experiment and counted as positive transformation.

The neutralizing properties of sera were tested by pre-incubating for 1 hour under slight agitation aliquotes of the Epstein-Barr Virus containing B95-8 cell supernatant with 20 µl of test serum including the respective preimmunization serum as control in a replicate test before the supernatant was allowed to adsorb to the umbilical cord blood lymphocytes. After removing the inoculum from the cells after 2 hours the maintenance medium (RPMI1640 supplemented with 10 % FCS) was supplemented with 5 % of the respective sera under test for neutralizing activity. The following results were obtained:

PBS (control)	Virus	Virus	Virus	Virus	Virus
(no virus)	+	+	+	+	+
	EBV negative human serum	EBV positive human serum pool	rabbit preserum	rabbit immune serum 1	rabbit immune serum 2
no colonies	colonies	no colonies	colonies	no colonies	no colonies

Example 20

Cloning and expression of antigenic fragments of the virus-capsid protein p150

The coding sequence of the diagnostically relevant protein p 150 (Virus capsid antigen VCA (see Example 1, Fig. 28D) was examined for antigenic sites and subcloned for the expression as ß-galactosidase fusion proteins. The N-terminal region which is expected to encode an antigenic site was obtained by digestion of the Charon 4A phage EB 69-79 (G.N. Buell, D. Reisman, C. Kintner, G. Crouse, and B. Sugden, "Cloning overlapping DNA fragments from the B95-8 strain of Epstein-Barr virus (ATCC CRL 1612) reveals a site of homology to the internal repetition", Journal of Virology 40, 977-982 (1981)) with BamHI and a resulting 1176bp fragment was cloned into the BamHI site of pUC12. From a resulting plasmid with the insertion in the proper orientation a 580bp fragment was excised with XhoI/SalI. The SalI site derives from the pUC12 linker, the XhoI site is located 33bp upstream from the start of p150. This fragment was inserted into pUC8 digested with SalI (SalI and XhoI share the same sticky end sequence). The resulting clones were screened to have the p150 start codon next to the BamHI site. From a proper clone the p150 encoding region was cut out with BamHI and HindIII and cloned into pUR290 digested with BamHI and HindIII (pUR290CXH580). The expression of the ß-Gal::p150 fusion protein from this clone is shown in Figure 29. Its ability to react very well with a NPC serum pool can be taken from Figure 30.

Further, p150::ß-gal fusion contructs were obtained accordingly. For example the subclones pUR290DBX320, pUR292DBB180, pUR290DTT 700, pURDTT740, pUR290DTP680, pUR288DPP320 which are indicated in Figure 31. From the designation of the subclones, the vector used can be taken,e.g. for the construction of subclone pUR290DBX320 the vector pUR290 was used.

From Figure 30 the restriction enzyme sites used for subcloning can also be taken. All clones with the exception of pURDBB180 were constructed by subcloning the desired fragments into pUC8 or pUC12 (see supra) to obtain BamHI and HindIII sites suitable for the cloning to pUR vectors (see supra). pUR292DBB180 was derived by insertion of the 180bp BglII-BglII fragment (see Fig.31) into pUR292 linearized with BamHI. Figures 29 and 30 show their expression and antigenicity.

The β-gal::p150 fusion protein encoded by pUR290CXH580 and purified according to example 4 reacts in the ELISA test as an EBV specific antigen indicating its applicability in diagnosis. Stable expression was also obtained with the N-terminal fragment of p150 by inserting the 580bp fragment (used for the construction of pUR290CSH580) into pUC18 (DSM 3424) using the BamHI and HindIII site. The resulting clone pUC18CXH580 expresses a stable and antigenic protein of about 25kD in size.

The following deposited plasmids, host bacteria and cell lines were used for the purpose of the present invention. The deposition was affected according to the Budapest treaty

m.o.	depository	deposition number
B 95.8	ATCC	CRL 1612
E.coli K12 JM83	ATCC	35607
E.coli K12 BMH71-18	DSM	3413
E.coli K12 JM109	DSM	3423
pUC8	DSM	3420
pUC9	DSM	3421
pUC12	DSM	3422
pUC19	DSM	3425
pUR288	DSM	3415
pUR290	DSM	3417
pUR291	DSM	3418
pUCARG680	DSM	3408
pUC18	DSM	3424

While we have hereinbefore presented a number of embodiments of this invention, it is apparent that our constructions can be alterd to provide other embodiments which utilize DNA sequences of the EBV genome coding for EBV-related antigens and for producing recombinant DNA molecules. It is obvious to those skilled in the art that other DNA sequences may also be used, which are related to said DNA sequences and which may be derived from other EBV serotypes. The EBV is easily obtainable from known natural sources, e. g. from the saliva of infected patients.

It is obvious that for obtaining biologically comparable results other suitable vector/host systems can be used. The invention ist not limited to host/vector systems presently available.

Claims:

1. A DNA sequence of the EBV-genome, characterized in that it corresponds to at least a part of an EBV-related antigenic protein having an aminoacid sequence as shown in Figures 3, 17, and 28.

2. A DNA sequence according to claim 1, characterized in that it corresponds to at least a part of protein p150, p143, p138, p110, p105, p90, p80, p54 or gp250/350.

3. A DNA sequence according to claims 1 or 2, characterized in that it contains additionally the respective regulatory sequences in the 5' and 3' flanks.

4. A DNA sequence hybridizing to a DNA sequence according to anyone of claims 1 to 3 from whatever source obtained including natural, synthetic or semisynthetic sources, which is related by mutations, including nucleotide substitutions, nucleotide deletions, nucleotide insertions and inversions of nucleotide stretches to a DNA sequence according to claims 1 to 3 and which encodes at least a part of a protein according to claim 1.

5. A DNA sequence according to claim 2, characterized in that it is inserted in the recombinant plasmid pUC6130, pUC635, pUCP400, pUCP380, pUCP600, pUCP210, pUCP750, pUCP540, pUCHP, pUC924, pMF924, pKK378, pUR600, pUR540, pUCARG680 or pUCARG1140.

6. A DNA sequence according to claim 2, characterized in that it is inserted in the recombinant plasmid pUCLP1.9, pURLP1.9, pUC19LEP600, pUC19LXP390, pURLXP390, pUCARG1230, pUCLEP600, pUCLXP390 and pURLEP600.

7. A DNA sequence according to claim 2, characterized in that it is inserted in the recombinant plasmid pUR290CXH580, pUR290DBX320, pUR292DBB180, pUR290DTT700, pURDTT740, pUR290DTP680 or pUR288DPP320.

8. A DNA sequence characterized in that in contains in reading frame at least two regions of a DNA sequence of anyone of claims 1 to 4 derived from a single EBV genome.

9. A DNA sequence according to claim 8, characterized in that it contains in reading frame at least two regions of a DNA sequence of anyone of claims 1 to 4 derived from different EBV genomes.

10. A DNA sequence according to anyone of claims 1 to 9, characterized in that it contains at its 3' end three to fifteen arginine codons positioned in the correct reading frame followed by at least one stop codon.

11. A DNA sequence according to anyone of claims 1 to 6, characterized in that it contains at its 5' end an oligonucleotide encoding an oligopeptide which serves in the resulting polypeptide as a cleavage site for a sequence specific protease or which is cleavable by acid treatment with an acid such as formic acid.

12. A recombinant DNA molecule for cloning, characterized in that it contains a DNA sequence according to anyone of claims 1 to 11.

13. A recombinant DNA molecule for expression, characterized in that it contains a DNA sequence according to anyone of the claims 1 to 11 that is operatively linked to an expresssion control sequence.

14. A recombinant DNA molecule according to claim 13, characterized in that the expression control sequence is selected from the group of the E. coli λpromoter system, the E. coli lac-system, the E. coli ß-lactamase system, the E. coli trp-system, the E. coli lipoprotein promoter, yeasts and other eukaryotic expression control sequences.

15. Vector carrying a part of the p138 encoding DNA sequence the encoded protein of which stabilizes in a fusion protein a protein encoded by a DNA sequence ligated to its 3'-end and carrying a DNA sequence encoding three to fifteen arginine residues followed by at least one stop codon which after insertion of the second DNA sequence is positioned at the 3'-end of this second sequence in the correct reading frame.

16. Vector according to claim 15 which is pUCARG601.

17. A host, characterized in that it is transformed by at least one recombinant DNA molecule according to anyone of claims 12 to 14.

18. A host according to claim 17 selected from the group consisting of strains of E. coli, other bacteria, yeasts, other fungi, animal and human cells.

19. A protein having EBV-related antigenic determinants suitable for diagnosis and therapy of EBV-related diseases, characterized in that it is encoded by a DNA sequence according to anyone of claims 1 to 11.

20. A polyantigen having at least two EBV-related antigenic determinants suitable for diagnosis and therapy of EBV-related diseases, characterized in that it is encoded by a DNA sequence according to anyone of claims 8 and 9.

21. A fusion protein, characterized in that it contains a protein according to claims 19 or 20.

22. A diagnostic composition for the detection of anti-EBV-antibodies, containing at least one protein according to anyone of claims 19 to 21 in an amount sufficient to bind said anti-EBV-antibodies in a sample.

23. A diagnostic composition for the detection of EBV-related diseases, containing at least one DNA sequence according to anyone of claims 1 to 11 in an amount sufficient for hybridization to an EBV-related DNA sequence in a sample.

24. A pharmaceutical composition containing at least one protein according to anyone of claims 19 to 21 in an amount sufficient for stimulating in humans the production of antibodies to EBV and a pharmaceutically acceptable carrier or diluent.

25. A method of preventing EBV infection or therapy of EBV-related diseases comprising administering to a human being the pharmaceutical composition according to claim 24 in an amount sufficient to induce or to modulate an immunoresponse.

Figure 1: Autoradiography of an immunoprecipitation of EBV-specific sera derived from patients suffering from mononucleosis and NPC.

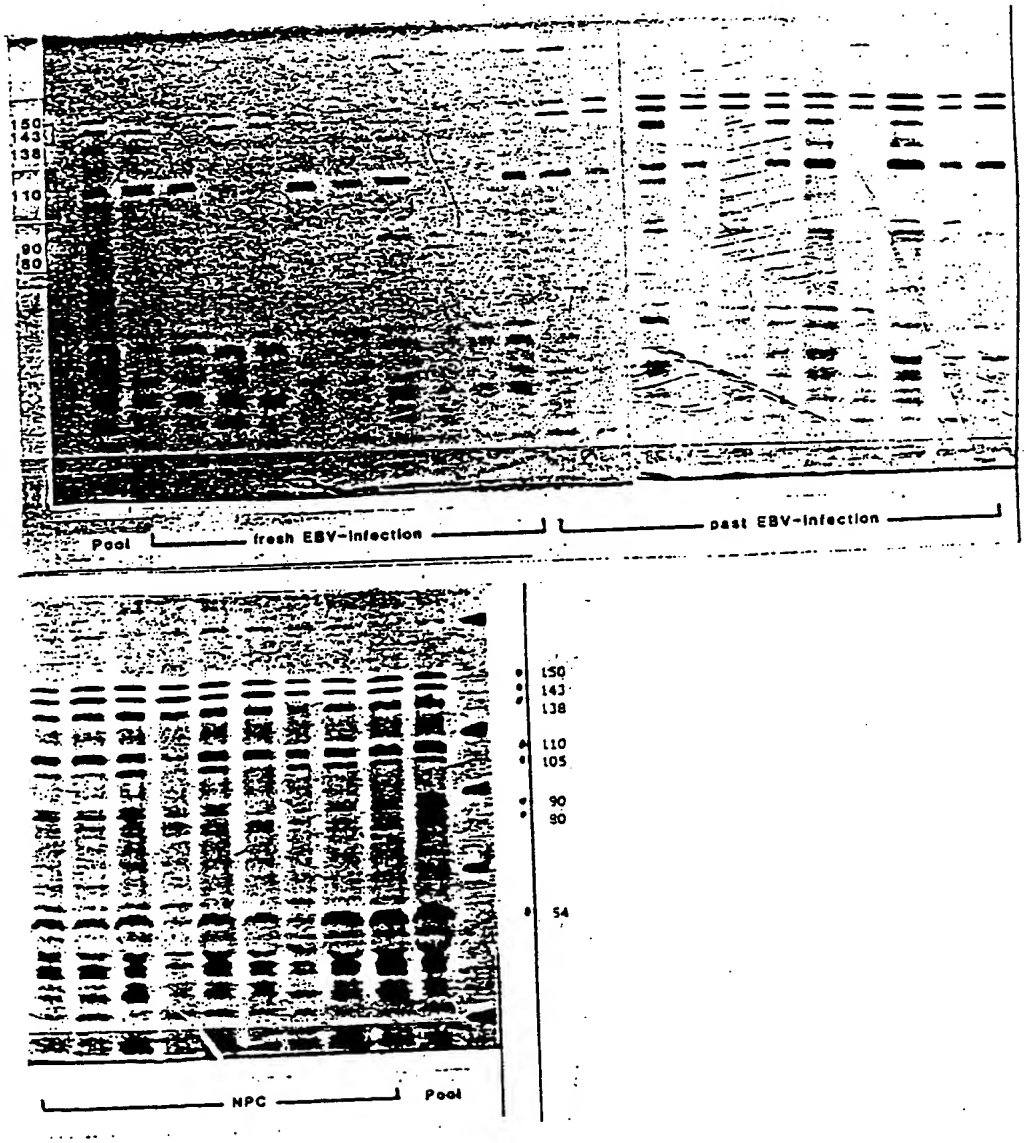
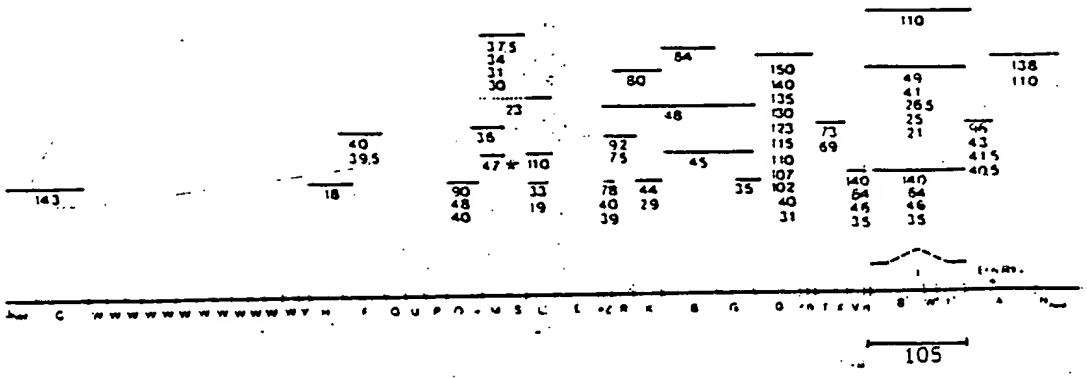


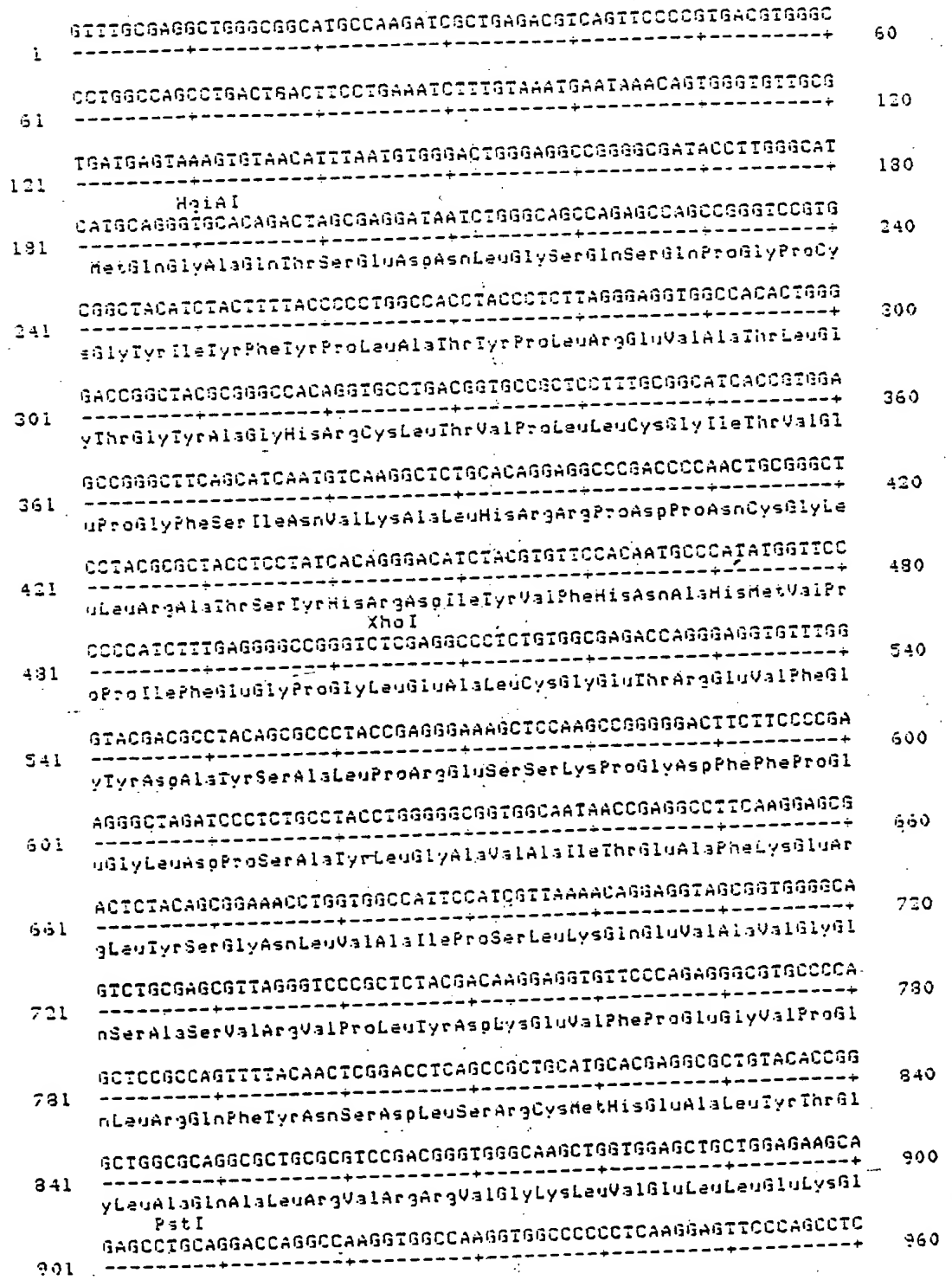
Figure 2

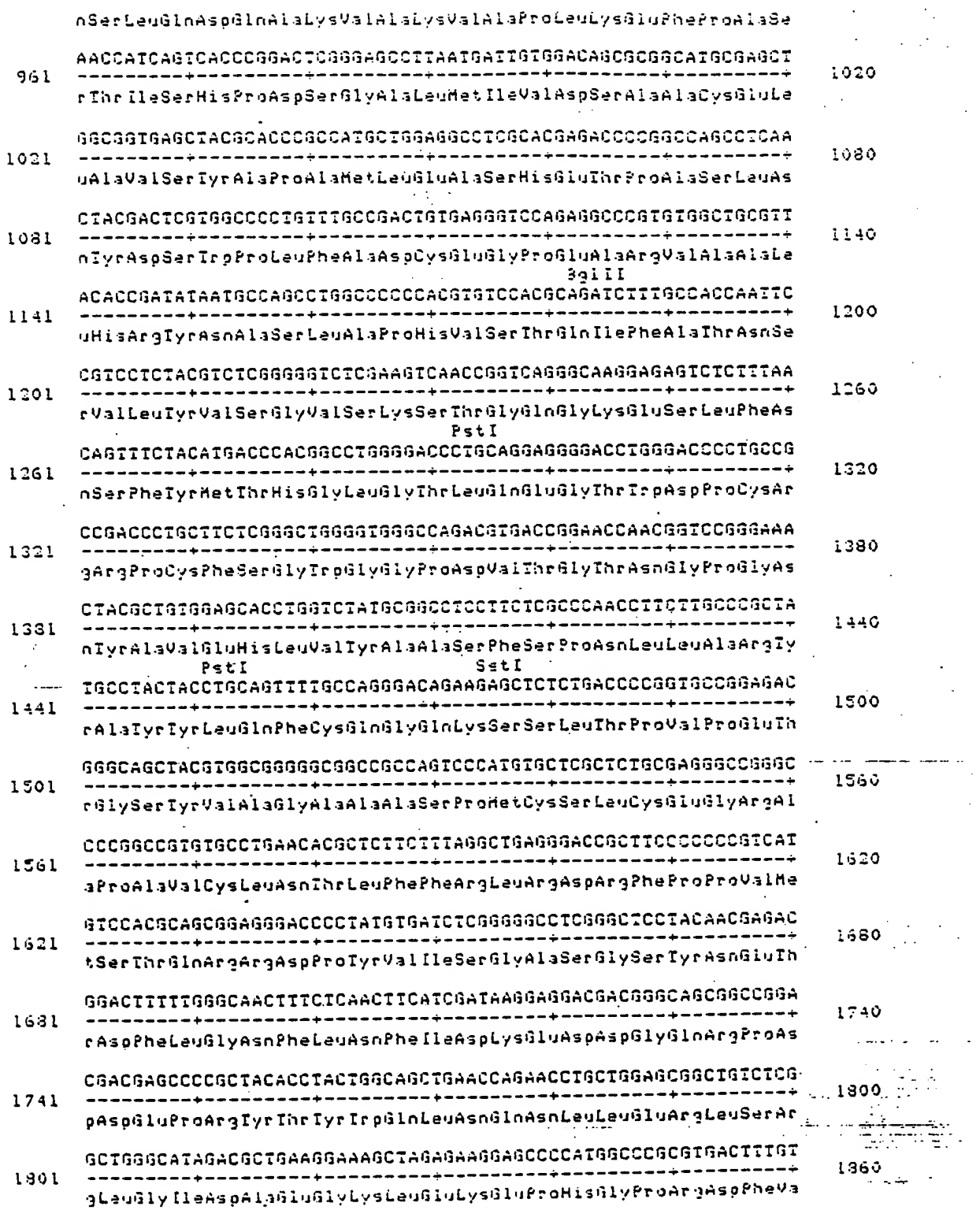
Mapping of mRNA's relative to the EBV B95-8 genome.

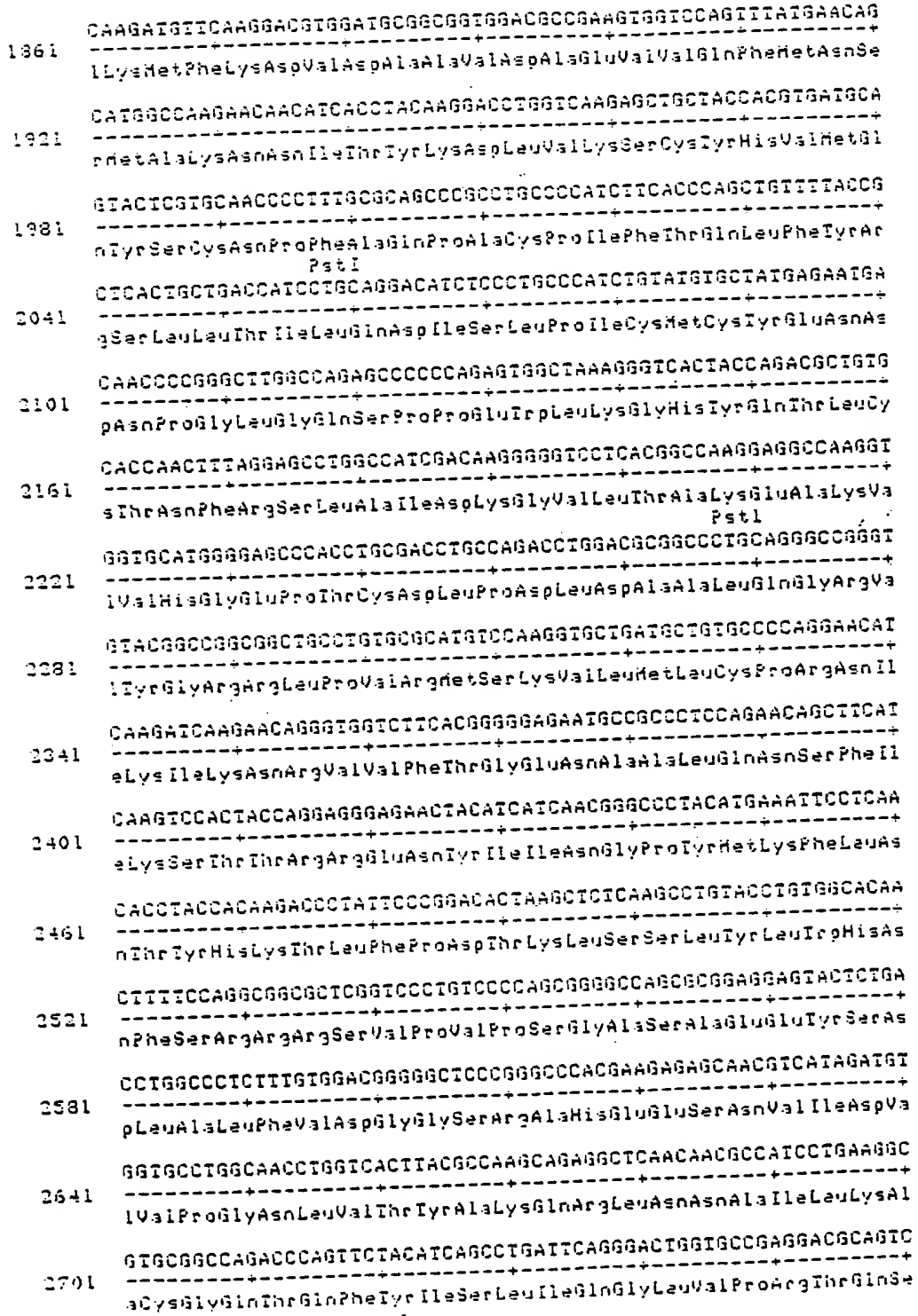


47* from translation in vitro correlates with p54 from in vivo labelling

Figure 3




nSerLeuGlnAspGlnAlaLysValAlaLysValAlaProLeuLysGluPheProAlaSe
AACCATCAGTCACCCGGACTCGGGAGCCTTAATGATTGTGGACAGCGCGGCATGCGAGCT
961
1020
rThrIleSerHisProAspSerGlyAlaLeuMetIleValAspSerAlaAlaCysGluLe
GGCGGTGAGCTACGCACCCGCCATGCTGGAGGCCTCGCACGAGACCCCGGCCAGCCTCAA
1021
1080
uAlaValSerTyrAlaProAlaMetLeuGluAlaSerHisGluThrProAlaSerLeuAs
CTACGACTCGTGGCCCCTGTTTGCCGACTGTGAGGGTCCAGAGGCCCGTGTGGCTGCGTT
1081
1140
nTyrAspSerTrpProLeuPheAlaAspCysGluGlyProGluAlaArgValAlaAlaLe
BglII
ACACCGATATAATGCCAGCCTGGCCCCCCACGTGTCCACGCAGATCTTTGCCACCAATTC
1141
1200
uHisArgTyrAsnAlaSerLeuAlaProHisValSerThrGlnIlePheAlaThrAsnSe
CGTCCTCTACGTCTCGGGGGTCTCGAAGTCAACCGGTCAGGGCAAGGAGAGTCTCTTTAA
1201
1260
rValLeuTyrValSerGlyValSerLysSerThrGlyGlnGlyLysGluSerLeuPheAs
PstI
CAGTTTCTACATGACCCACGGCCTGGGGACCCTGCAGGAGGGGACCTGGGACCCCTGCCG
1261
1320
nSerPheTyrMetThrHisGlyLeuGlyThrLeuGlnGluGlyThrTrpAspProCysAr
CCGACCCTGCTTCTCGGGCTGGGGTGGGCCAGACGTGACCGGAACCAACGGTCCGGGAAA
1321
1380
gArgProCysPheSerGlyTrpGlyGlyProAspValThrGlyThrAsnGlyProGlyAs
CTACGCTGTGGAGCACCTGGTCTATGCGGCCTCCTTCTCGCCCAACCTTCTTGCCCGCTA
1381
1440
nTyrAlaValGluHisLeuValTyrAlaAlaSerPheSerProAsnLeuLeuAlaArgTy
PstI
SstI
TGCCTACTACCTGCAGTTTTGCCAGGGACAGAAGAGCTCTCTGACCCCGGTGCCGGAGAC
1441
1500
rAlaTyrTyrLeuGlnPheCysGlnGlyGlnLysSerSerLeuThrProValProGluTh
GGGCAGCTACGTGGCGGGGGCGGCCGCCAGTCCCATGTGCTCGCTCTGCGAGGGCCGGGC
1501
1560
rGlySerTyrValAlaGlyAlaAlaAlaSerProMetCysSerLeuCysGluGlyArgAl
CCCGGCCGTGTGCCTGAACACGCTCTTCTTTAGGCTGAGGGACCGCTTCCCCCCCGTCAT
1561
1620
aProAlaValCysLeuAsnThrLeuPhePheArgLeuArgAspArgPheProProValMe
GTCCACGCAGCGGAGGGACCCCTATGTGATCTCGGGGGCCTCGGGCTCCTACAACGAGAC
1621
1680
tSerThrGlnArgArgAspProTyrValIleSerGlyAlaSerGlySerTyrAsnGluTh
GGACTTTTTGGGCAACTTTCTCAACTTCATCGATAAGGAGGACGACGGGCAGCGGCCGGA
1681
1740
rAspPheLeuGlyAsnPheLeuAsnPheIleAspLysGluAspAspGlyGlnArgProAs
CGACGAGCCCCGCTACACCTACTGGCAGCTGAACCAGAACCTGCTGGAGCGGCTGTCTCG
1741
1800
pAspGluProArgTyrThrTyrTrpGlnLeuAsnGlnAsnLeuLeuGluArgLeuSerAr
GCTGGGCATAGACGCTGAAGGAAAGCTAGAGAAGGAGCCCCATGGCCCGCGTGACTTTGT
1801
1860
gLeuGlyIleAspAlaGluGlyLysLeuGluLysGluProHisGlyProArgAspPheVa


1861 CAAGATGTTCAAGGACGTGGATGCGGCGGTGGACGCCGAAGTGGTCCAGTTTATGAACAG 1920
lLysMetPheLysAspValAspAlaAlaValAspAlaGluValValGlnPheMetAsnSe
1921 CATGGCCAAGAACAACATCACCTACAAGGACCTGGTCAAGAGCTGCTACCACGTGATGCA 1980
rMetAlaLysAsnAsnIleThrTyrLysAspLeuValLysSerCysTyrHisValMetGl
1981 GTACTCGTGCAACCCCTTTGCGCAGCCCGCCTGCCCCATCTTCACCCAGCTGTTTTACCG 2040
nTyrSerCysAsnProPheAlaGlnProAlaCysProIlePheThrGlnLeuPheTyrAr
PstI
2041 CTCACTGCTGACCATCCTGCAGGACATCTCCCTGCCCATCTGTATGTGCTATGAGAATGA 2100
gSerLeuLeuThrIleLeuGlnAspIleSerLeuProIleCysMetCysTyrGluAsnAs
2101 CAACCCCGGGCTTGGCCAGAGCCCCCCAGAGTGGCTAAAGGGTCACTACCAGACGCTGTG 2160
pAsnProGlyLeuGlyGlnSerProProGluTrpLeuLysGlyHisTyrGlnThrLeuCy
2161 CACCAACTTTAGGAGCCTGGCCATCGACAAGGGGGTCCTCACGGCCAAGGAGGCCAAGGT 2220
sThrAsnPheArgSerLeuAlaIleAspLysGlyValLeuThrAlaLysGluAlaLysVa
PstI
2221 GGTGCATGGGGAGCCCACCTGCGACCTGCCAGACCTGGACGCGGCCCTGCAGGGCCGGGT 2280
lValHisGlyGluProThrCysAspLeuProAspLeuAspAlaAlaLeuGlnGlyArgVa
2281 GTACGGCCGGCGGCTGCCTGTGCGCATGTCCAAGGTGCTGATGCTGTGCCCCAGGAACAT 2340
lTyrGlyArgArgLeuProValArgMetSerLysValLeuMetLeuCysProArgAsnIl
2341 CAAGATCAAGAACAGGGTGGTCTTCACGGGGGAGAATGCCGCCCTCCAGAACAGCTTCAT 2400
eLysIleLysAsnArgValValPheThrGlyGluAsnAlaAlaLeuGlnAsnSerPheIl
2401 CAAGTCCACTACCAGGAGGGAGAACTACATCATCAACGGGCCCTACATGAAAATTCCTCAA 2460
eLysSerThrThrArgArgGluAsnTyrIleIleAsnGlyProTyrMetLysPheLeuAs
2461 CACCTACCACAAGACCCTATTCCCGGACACTAAGCTCTCAAGCCTGTACCTGTGGCACAA 2520
nThrTyrHisLysThrLeuPheProAspThrLysLeuSerSerLeuTyrLeuTrpHisAs
2521 CTTTTCCAGGCGGCGCTCGGTCCCTGTCCCCAGCGGGGCCAGCGCGGAGGAGTACTCTGA 2580
nPheSerArgArgArgSerValProValProSerGlyAlaSerAlaGluGluTyrSerAs
2581 CCTGGCCCTCTTTGTGGACGGGGGCTCCCGGGCCCACGAAGAGAGCAACGTCATAGATGT 2640
pLeuAlaLeuPheValAspGlyGlySerArgAlaHisGluGluSerAsnValIleAspVa
2641 GGTGCCTGGCAACCTGGTCACTTACGCCAAGCAGAGGCTCAACAACGCCATCCTGAAGGC 2700
lValProGlyAsnLeuValThrTyrAlaLysGlnArgLeuAsnAsnAlaIleLeuLysAl
2701 GTGCGGCCAGACCCAGTTCTACATCAGCCTGATTCAGGGACTGGTGCCGAGGACGCAGTC 2760
aCysGlyGlnThrGlnPheTyrIleSerLeuIleGlnGlyLeuValProArgThrGlnSe

```
                GGTGCCCGCCCGTGACTACCCCCACGTACTGGGCACGCGGGCGGTGGAGTCGGCAGCGGC
2761            ---------+---------+---------+---------+---------+---------+  3820
                rValProAlaArgAspTyrProHisValLeuGlyThrArgAlaValGluSerAlaAlaAl

                CTACGCGGAGGCCACCTCCTCCCTTACTGCGACCACGGTGGTCTGCGCGGCCACAGACTG
3821            ---------+---------+---------+---------+---------+---------+  3880
                aTyrAlaGluAlaThrSerSerLeuThrAlaThrThrValValCysAlaAlaThrAspCy

                TCTTAGCCAGGTCTGCAAGGCCCGTCCGGTTGTCACGCTGCCAGTGACCATCAACAAGTA
3881            ---------+---------+---------+---------+---------+---------+  3940
                sLeuSerGlnValCysLysAlaArgProValValThrLeuProValThrIleAsnLysTy

                CACGGGGGTCAACGGCAACAACCAGATATTCCAGGCCGGGAACCTGGGATACTTTATGGG
3941            ---------+---------+---------+---------+---------+---------+  4000
                rThrGlyValAsnGlyAsnAsnGlnIlePheGlnAlaGlyAsnLeuGlyTyrPheMetGl
                                        PstI
                CCGGGGCGTGGACAGGAACCTGCTGCAGGCCCCCGGGGCTGGGCTGCGCAAGCAGGCCGG
4001            ---------+---------+---------+---------+---------+---------+  4060
                yArgGlyValAspArgAsnLeuLeuGlnAlaProGlyAlaGlyLeuArgLysGlnAlaGl

                GGGCTCTTCCATGCGGAAGAAGTTTGTCTTTGCCACCCCCACCCTAGGGTTGACCGTGAA
4061            ---------+---------+---------+---------+---------+---------+  4120
                yGlySerSerMetArgLysLysPheValPheAlaThrProThrLeuGlyLeuThrValLy

                GCGCCGGACCCAAGCCGCGACCACATATGAGATTGAGAACATCAGGGCTGGCCTGGAGGC
4121            ---------+---------+---------+---------+---------+---------+  4180
                sArgArgThrGlnAlaAlaThrThrTyrGluIleGluAsnIleArgAlaGlyLeuGluAl

                CATTATATCACAAAAACAGGAGGAAGACTGTGTGTTTGATGTGGTGTGCAACCTTGTGGA
4181            ---------+---------+---------+---------+---------+---------+  4240
                aIleIleSerGlnLysGlnGluGluAspCysValPheAspValValCysAsnLeuValAs

                TGCCATGGGCGAGGCATGCGCCTCGCTGACTAGGGACGACGCGGAGTACTTATTGGGCCG
4241            ---------+---------+---------+---------+---------+---------+  4300
                pAlaMetGlyGluAlaCysAlaSerLeuThrArgAspAspAlaGluTyrLeuLeuGlyAr

                CTTCTCCGTCCTGGCGGACAGCGTCCTAGAAACCCTGGCGACCATTGCCTCCAGCGGGAT
4301            ---------+---------+---------+---------+---------+---------+  4360
                gPheSerValLeuAlaAspSerValLeuGluThrLeuAlaThrIleAlaSerSerGlyIl

                AGAGTGGACGGCGGAGGCCGCTCGGGACTTTCTGGAGGGAGTGTGGGGTGGGCCCGGGGC
4361            ---------+---------+---------+---------+---------+---------+  4420
                eGluTrpThrAlaGluAlaAlaArgAspPheLeuGluGlyValTrpGlyGlyProGlyAl

                AGCCCAGGACAACTTTATCAGCGTGGCCGAGCCGGTCAGCACCGCGTCGCAGGCCTCGGC
4421            ---------+---------+---------+---------+---------+---------+  4480
                aAlaGlnAspAsnPheIleSerValAlaGluProValSerThrAlaSerGlnAlaSerAl

                CGGGCTGCTGCTGGGTGGAGGAGGGCAGGGCTCCGGGGGCAGACGCAAGCGCCGTCTGGC
4481            ---------+---------+---------+---------+---------+---------+  4540
                aGlyLeuLeuLeuGlyGlyGlyGlyGlnGlySerGlyGlyArgArgLysArgArgLeuAl
                                    XhoI
                CACCGTTCTCCCCGGACTCGAGGTCTAGAGACCCCTGGGGCGGCGATGTCGGGGCTGCTG
4541            ---------+---------+---------+---------+---------+---------+  4600
                aThrValLeuProGlyLeuGluValEnd

                GCGGCGGCGTACAGCCAGGTGTACGCCCTGGCGGTTGAGCTGAGCGTGTGCACCCGGCTG
4601            ---------+---------+---------+---------+---------+---------+  4660

                GACCCCCGGAGTCTGGACGTGGCTGCGGTGGTGCGCAACGCCGGCCTGCTGGCCGAGCTG
4661            ---------+---------+---------+---------+---------+---------+  4720
```

```
      GAGGCCATCCTCCTTCCCCGTTTGAGACGGCAGAATGACCGTGCATGCAGCGCCCTGTCC
3721  ---------+---------+---------+---------+---------+---------+  3780
                                    XhoI
      CTGGAGCTGGTGCACCTGCTAGAGAACTCGAGAGAGGCCTCTGCCGCGCTGCTCGCCCCT
3781  ---------+---------+---------+---------+---------+---------+  3840

      GGTAGAAAGGG
3841  ----------+-  3851
```

Figure 4:
Restriction map of the plasmids pUC635 and pUC6130.

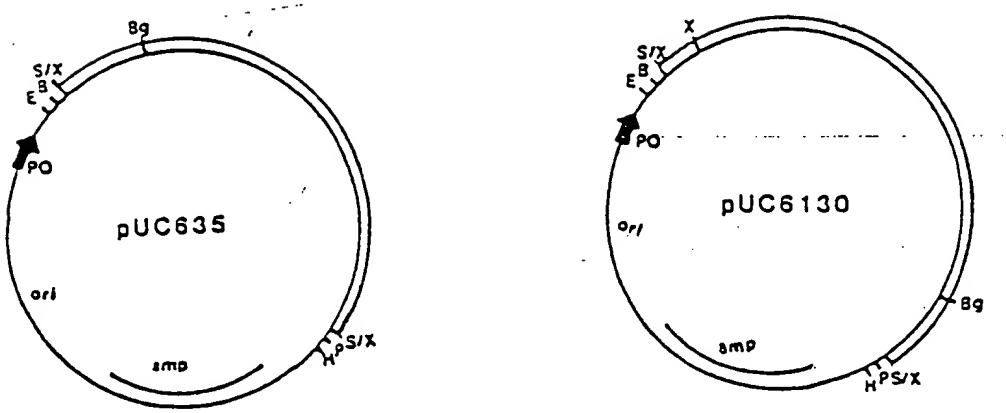


Figure 5:

Expression of the p138 fusion protein encoded by pUC635, pUC924, pMF924, and pKK378

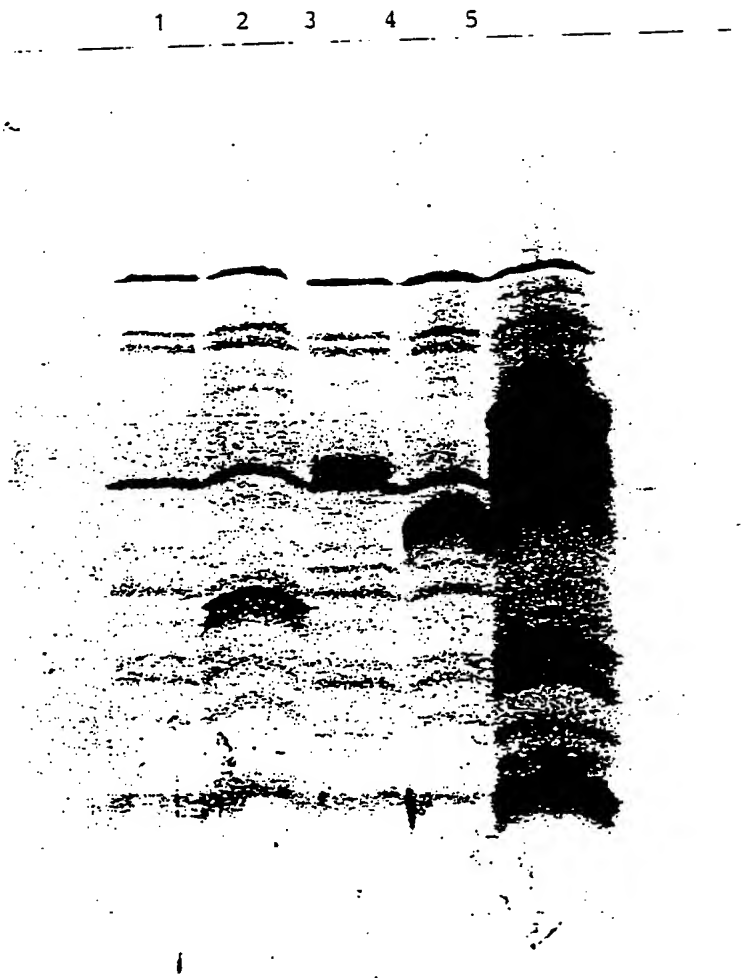


Figure 6:
Restriction map of the plasmid pUC924

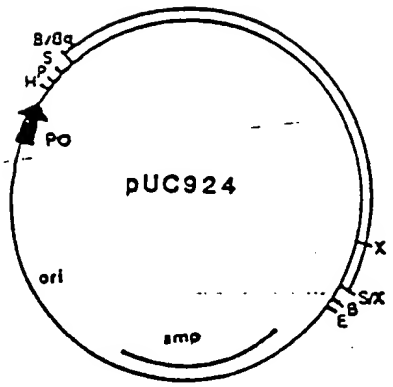


Figure 7:
Restriction map of the plasmid pMF924.

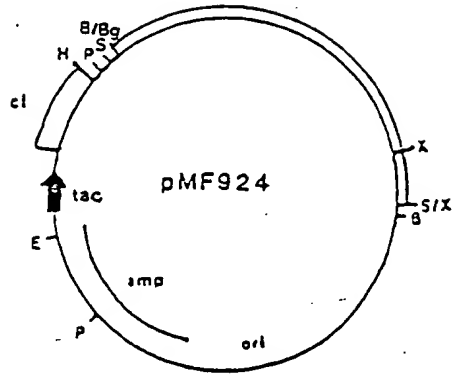


Figure 8:
Restriction map of the plasmid pKK378.

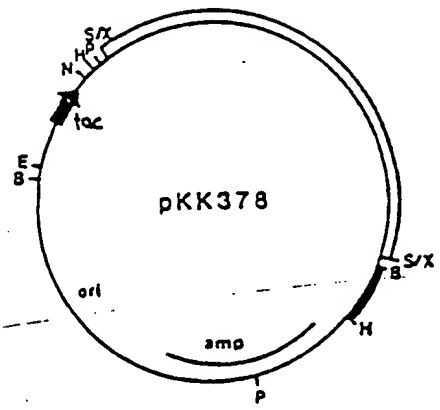


Figure 9:
Secondary structures of p138.

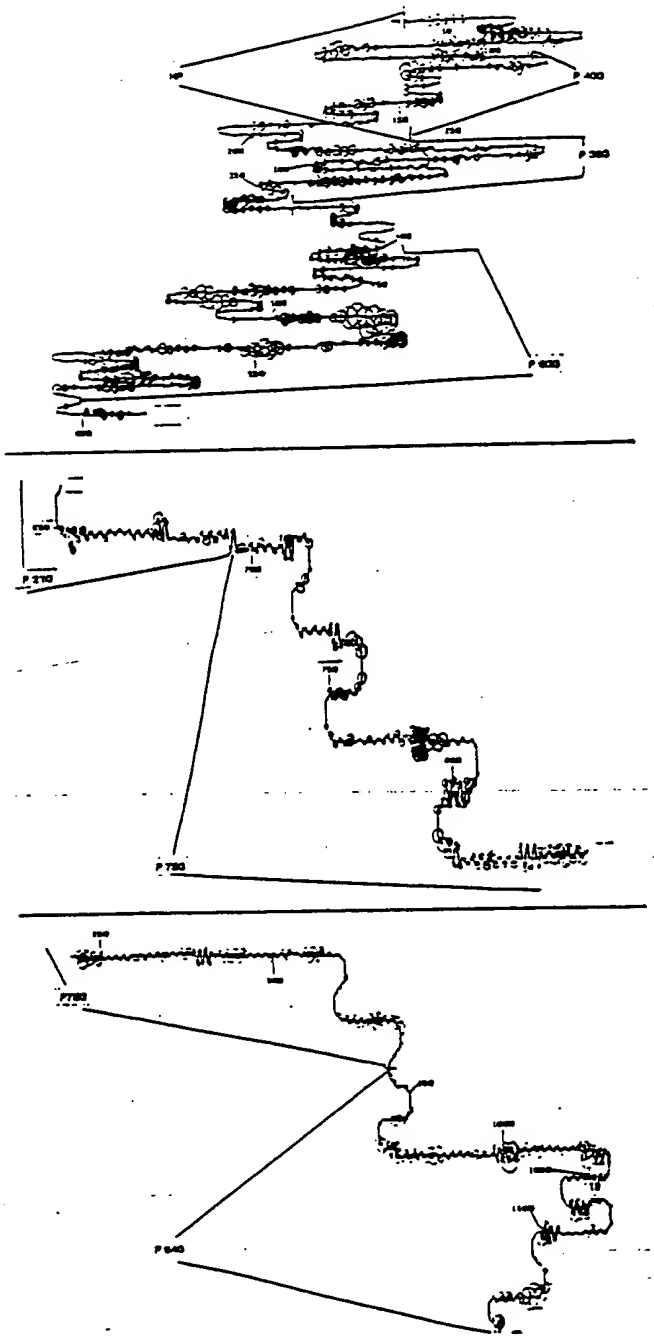
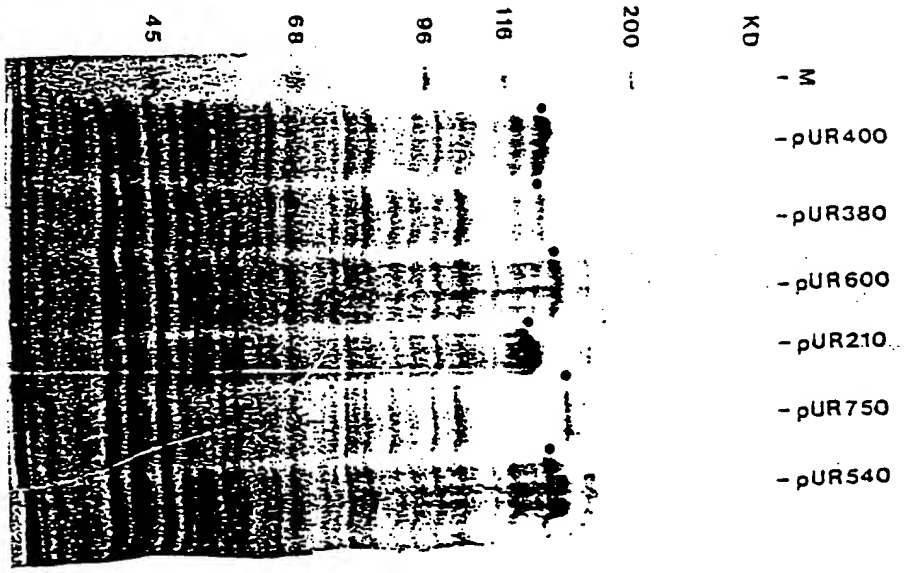
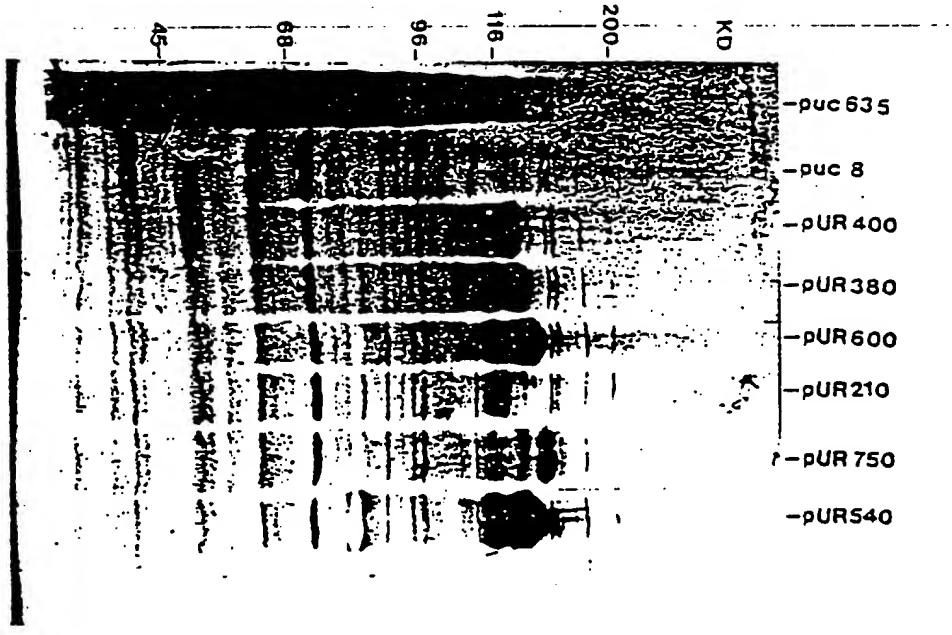


Figure 10:
Expression products of bacteria transformed with the pUR-carrying PstI fragments of p138.



A



B

Figure 11:
Expression products of bacteria transformed with the pUC subclones carrying PstI-fragments of p138.

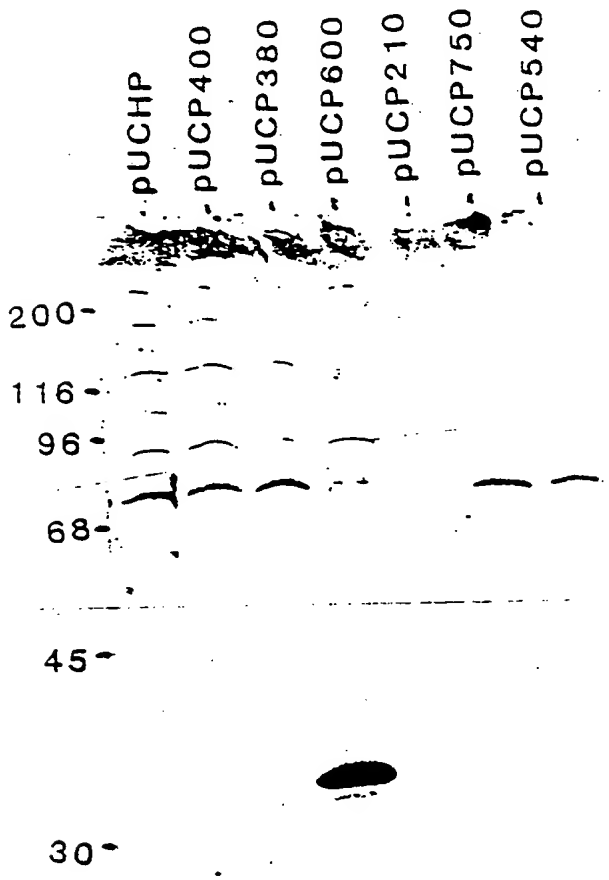
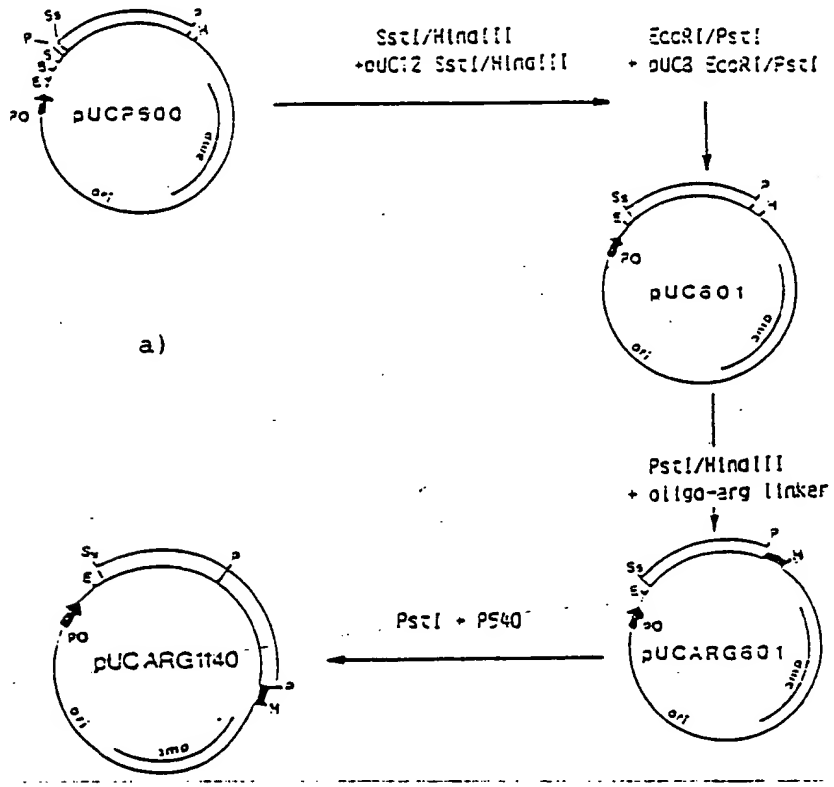


Figure 12:
Construction scheme for pUCARG1140 encoding both antigenic sites found by expression as β-gal fusion proteins



b)

PstI 5' ... 3' HindIII

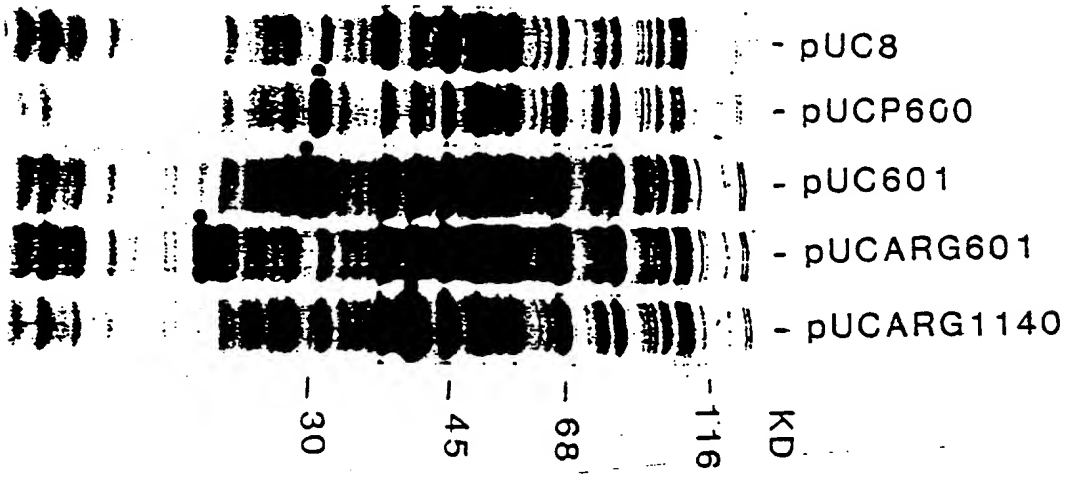
G CGT CGT CGT CGT CGT TGA TA

AC GTC GCA GCA GCA GCA GCA ACT ATT CGA

Arg Arg Arg Arg Arg stop stop

Figure 13:

IPTG-induced expression of the plasmids pUC600, pUC601, pUCARG601 and pUCARG1140 with pUC8 as a control



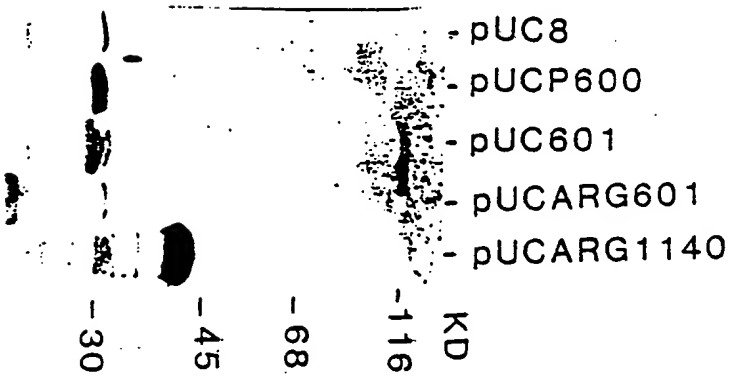


Figure 14:

Distribution and reactivity of the IgG and IgA antibodies of individual NPC-sera against the two epitopes detected in p138

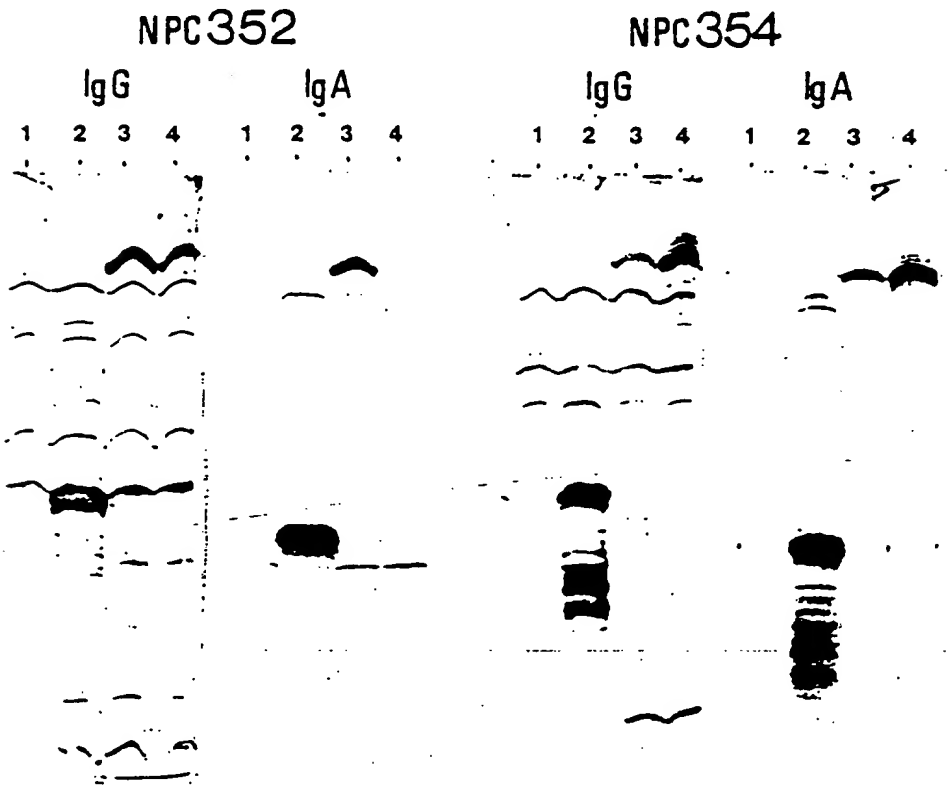


Figure 15:

ELISA test using the protein encoded by pUCARG1140 as antigen

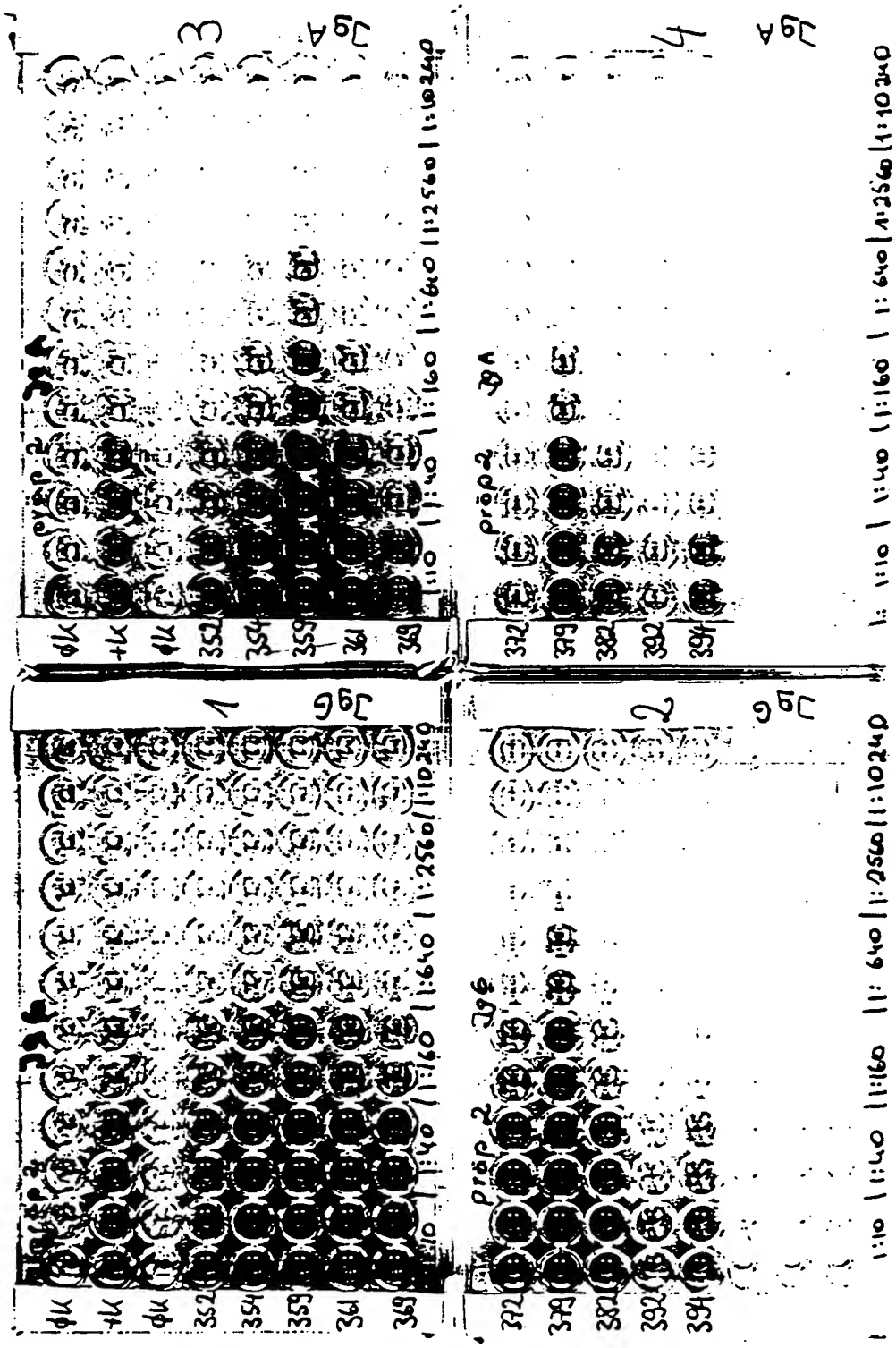


Figure 16:
Purification of proteins carrying oligo-arginine peptides at their carboxy-terminus.

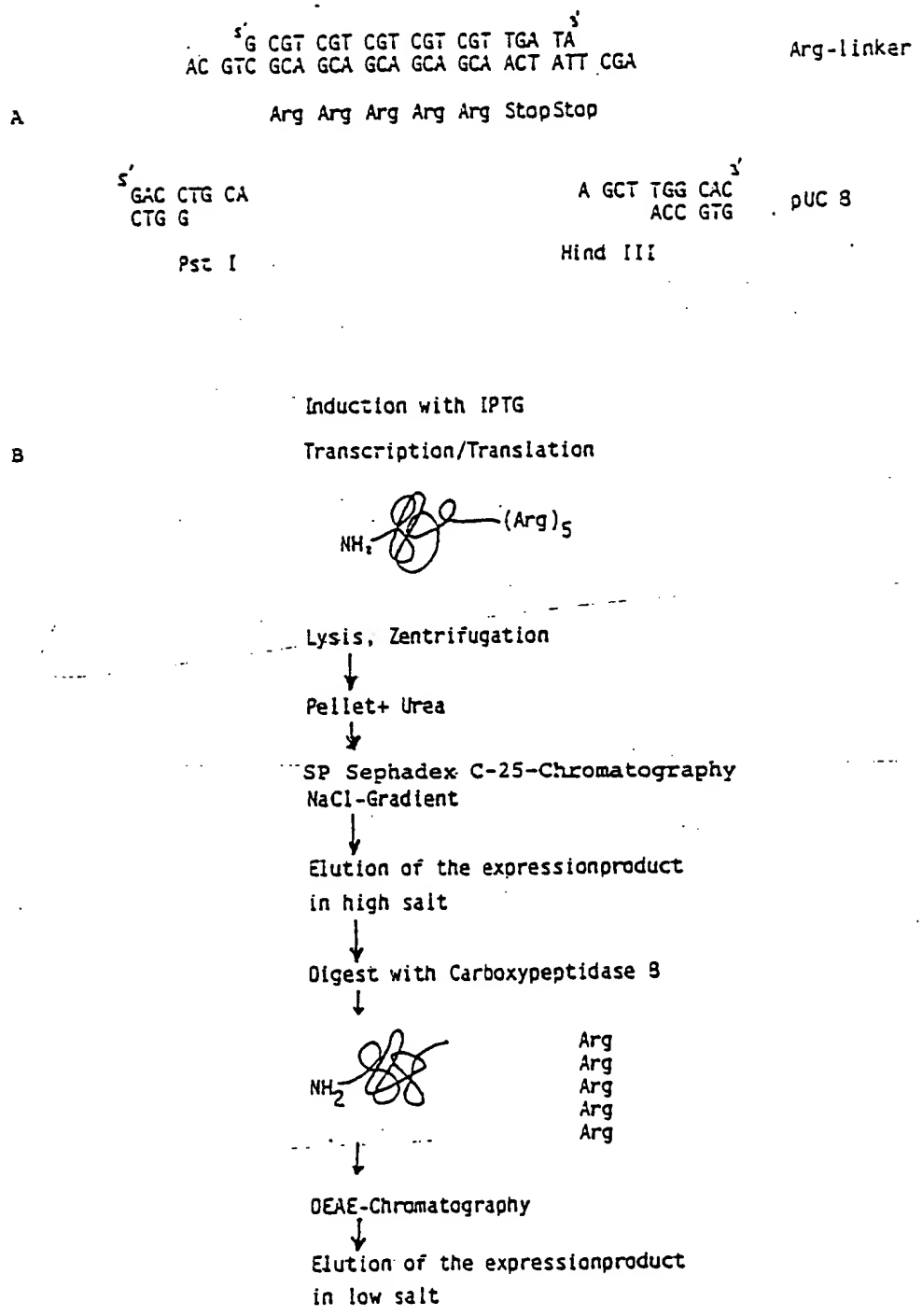
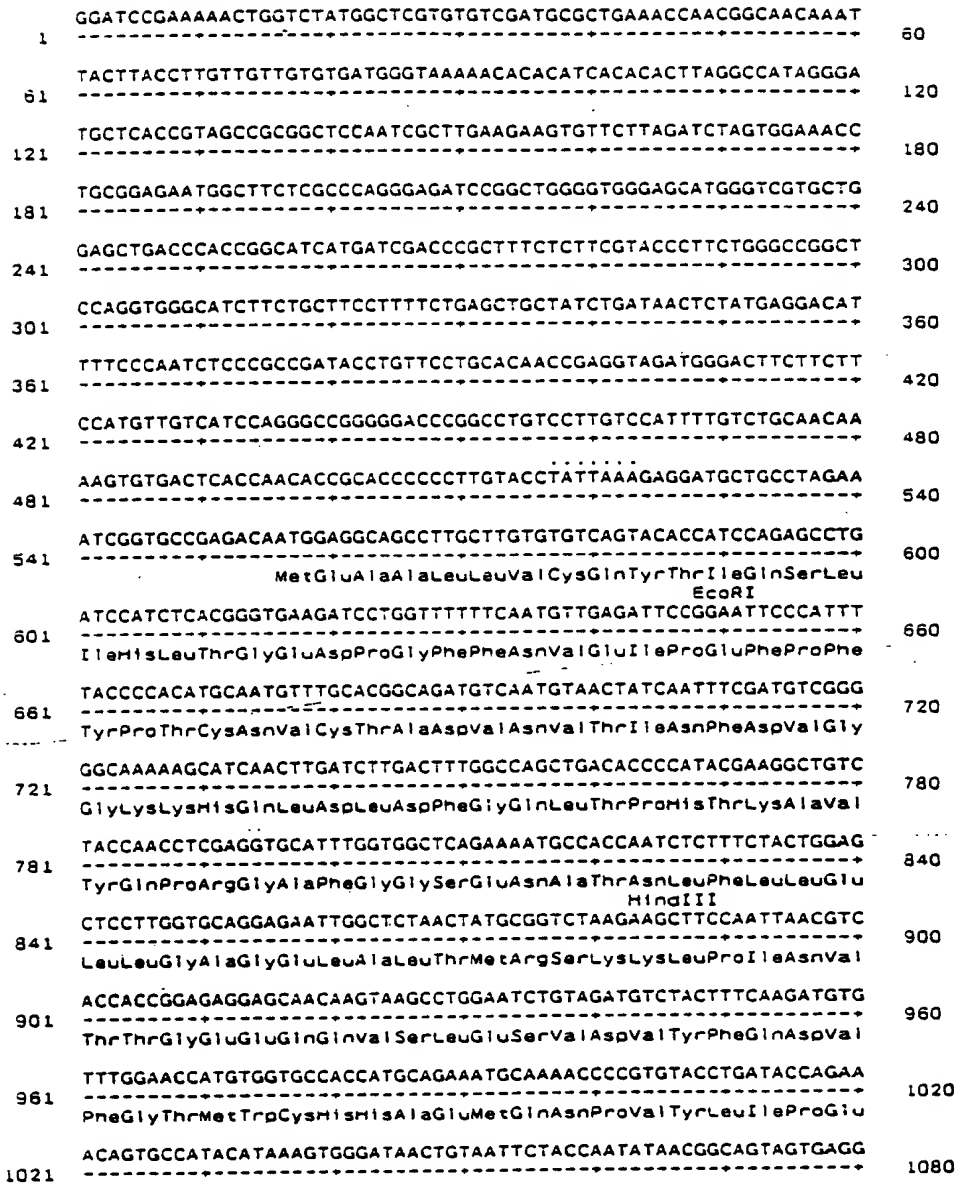
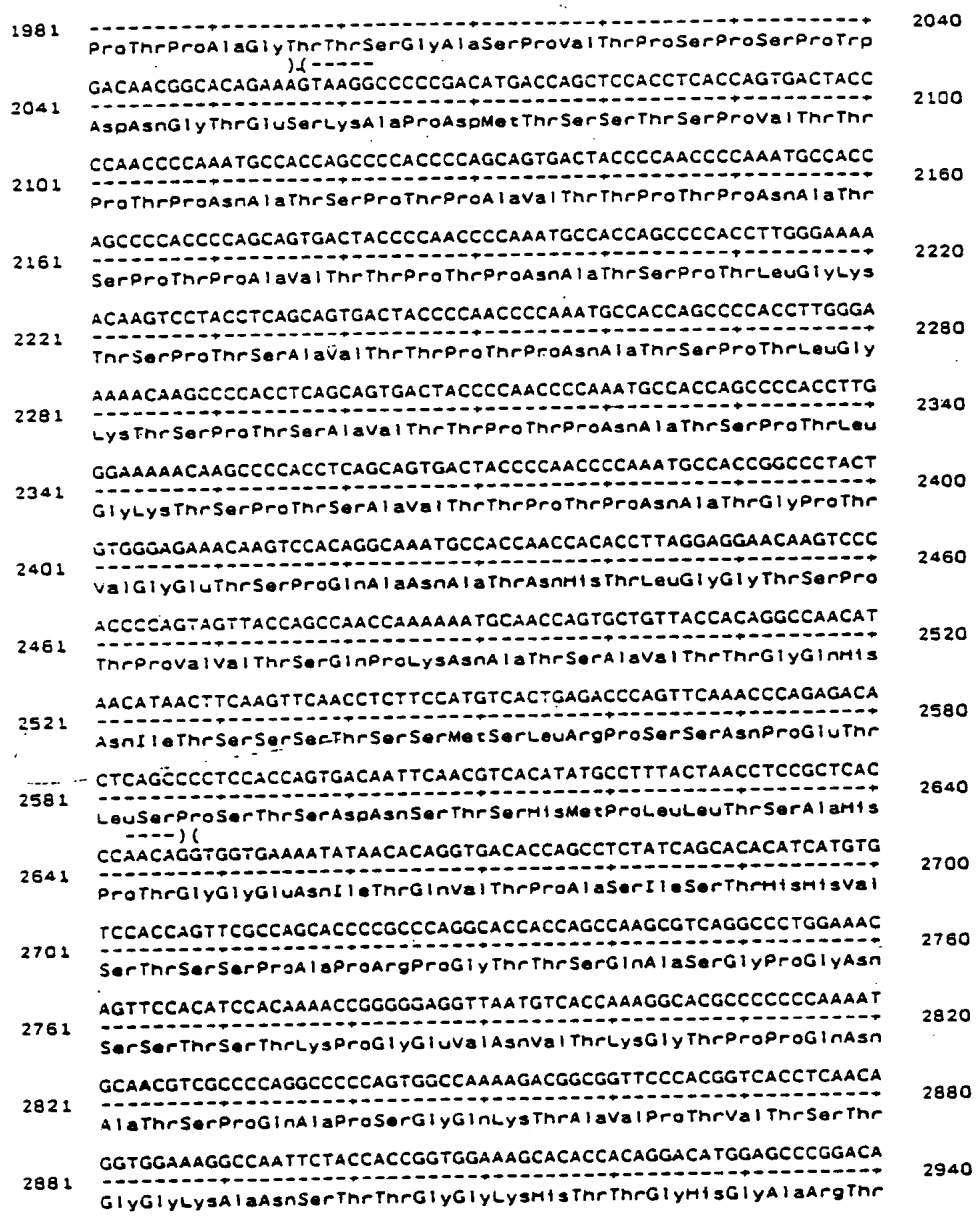


Figure 17


GGATCCGAAAAACTGGTCTATGGCTCGTGTGTCGATGCGCTGAAACCAACGGCAACAAAT
1
60
TACTTACCTTGTTGTTGTGTGATGGGTAAAAACACACATCACACACTTAGGCCATAGGGA
61
120
TGCTCACCGTAGCCGCGGCTCCAATCGCTTGAAGAAGTGTTCTTAGATCTAGTGGAAACC
121
180
TGCGGAGAATGGCTTCTCGCCCAGGGAGATCCGGCTGGGGTGGGAGCATGGGTCGTGCTG
181
240
GAGCTGACCCACCGGCATCATGATCGACCCGCTTTCTCTTCGTACCCTTCTGGGCCGGCT
241
300
CCAGGTGGGCATCTTCTGCTTCCTTTTCTGAGCTGCTATCTGATAACTCTATGAGGACAT
301
360
TTTCCCAATCTCCCGCCGATACCTGTTCCTGCACAACCGAGGTAGATGGGACTTCTTCTT
361
420
CCATGTTGTCATCCAGGGCCGGGGGACCCGGCCTGTCCTTGTCCATTTTGTCTGCAACAA
421
480
AAGTGTGACTCACCAACACCGCACCCCCCTTGTACCTATTAAAGAGGATGCTGCCTAGAA
481
540
ATCGGTGCCGAGACAATGGAGGCAGCCTTGCTTGTGTGTCAGTACACCATCCAGAGCCTG
541
600
MetGluAlaAlaLeuLeuValCysGlnTyrThrIleGlnSerLeu
EcoRI
ATCCATCTCACGGGTGAAGATCCTGGTTTTTTCAATGTTGAGATTCCGGAATTCCCATTT
601
660
IleHisLeuThrGlyGluAspProGlyPhePheAsnValGluIleProGluPheProPhe
TACCCCACATGCAATGTTTGCACGGCAGATGTCAATGTAACTATCAATTTCGATGTCGGG
661
720
TyrProThrCysAsnValCysThrAlaAspValAsnValThrIleAsnPheAspValGly
GGCAAAAAGCATCAACTTGATCTTGACTTTGGCCAGCTGACACCCCATACGAAGGCTGTC
721
780
GlyLysLysHisGlnLeuAspLeuAspPheGlyGlnLeuThrProHisThrLysAlaVal
TACCAACCTCGAGGTGCATTTGGTGGCTCAGAAAATGCCACCAATCTCTTTCTACTGGAG
781
840
TyrGlnProArgGlyAlaPheGlyGlySerGluAsnAlaThrAsnLeuPheLeuLeuGlu
HindIII
CTCCTTGGTGCAGGAGAATTGGCTCTAACTATGCGGTCTAAGAAGCTTCCAATTAACGTC
841
900
LeuLeuGlyAlaGlyGluLeuAlaLeuThrMetArgSerLysLysLeuProIleAsnVal
ACCACCGGAGAGGAGCAACAAGTAAGCCTGGAATCTGTAGATGTCTACTTTCAAGATGTG
901
960
ThrThrGlyGluGluGlnGlnValSerLeuGluSerValAspValTyrPheGlnAspVal
TTTGGAACCATGTGGTGCCACCATGCAGAAATGCAAAACCCCGTGTACCTGATACCAGAA
961
1020
PheGlyThrMetTrpCysHisHisAlaGluMetGlnAsnProValTyrLeuIleProGlu
ACAGTGCCATACATAAAGTGGGATAACTGTAATTCTACCAATATAACGGCAGTAGTGAGG
1021
1080

```
                ThrValProTyrIleLysTrpAspAsnCysAsnSerThrAsnIleThrAlaValValArg

                GCACAGGGGCTGGATGTCACGCTACCCTTAAGTTTGCCAACGTCAGCTCAAGACTCGAAT
1081            ----------+---------+---------+---------+---------+---------+ 1140
                AlaGlnGlyLeuAspValThrLeuProLeuSerLeuProThrSerAlaGlnAspSerAsn

                TTCAGCGTAAAAACAGAAATGCTCGGTAATGAGATAGATATTGAGTGTATTATGGAGGAT
1141            ----------+---------+---------+---------+---------+---------+ 1200
                PheSerValLysThrGluMetLeuGlyAsnGluIleAspIleGluCysIleMetGluAsp
                                                                  PstI
                GGCGAAATTTCACAAGTTCTGCCCGGAGACAACAAATTTAACATCACCTGCAGTGGATAC
1201            ----------+---------+---------+---------+---------+---------+ 1260
                GlyGluIleSerGlnValLeuProGlyAspAsnLysPheAsnIleThrCysSerGlyTyr
                                  EcoRI
                GAGAGCCATGTTCCCAGCGGCGGAATTCTCACATCAACGAGTCCCGTGGCCACCCCAATA
1261            ----------+---------+---------+---------+---------+---------+ 1320
                GluSerHisValProSerGlyGlyIleLeuThrSerThrSerProValAlaThrProIle

                CCTGGTACAGGGTATGCATACAGCCTGCGTCTGACACCACGTCCAGTGTCACGATTTCTT
1321            ----------+---------+---------+---------+---------+---------+ 1380
                ProGlyThrGlyTyrAlaTyrSerLeuArgLeuThrProArgProValSerArgPheLeu

                GGCAATAACAGTATCCTGTACGTGTTTTACTCTGGGAATGGACCGAAGGCGAGCGGGGGA
1381            ----------+---------+---------+---------+---------+---------+ 1440
                GlyAsnAsnSerIleLeuTyrValPheTyrSerGlyAsnGlyProLysAlaSerGlyGly

                GATTACTGCATTCAGTCCAACATTGTGTTCTCTGATGAGATTCCAGCTTCACAGGACATG
1441            ----------+---------+---------+---------+---------+---------+ 1500
                AspTyrCysIleGlnSerAsnIleValPheSerAspGluIleProAlaSerGlnAspMet

                CCGACAAACACCACAGACATCACATATGTGGGTGACAATGCTACCTATTCAGTGCCAATG
1501            ----------+---------+---------+---------+---------+---------+ 1560
                ProThrAsnThrThrAspIleThrTyrValGlyAspAsnAlaThrTyrSerValProMet

                GTCACTTCTGAGGACGCAAACTCGCCAAATGTTACAGTGACTGCCTTTTGGGCCTGGCCA
1561            ----------+---------+---------+---------+---------+---------+ 1620
                ValThrSerGluAspAlaAsnSerProAsnValThrValThrAlaPheTrpAlaTrpPro

                AACAACACTGAAACTGACTTTAAGTGCAAATGGACTCTCACCTCGGGGACACCTTCGGGT
1621            ----------+---------+---------+---------+---------+---------+ 1680
                AsnAsnThrGluThrAspPheLysCysLysTrpThrLeuThrSerGlyThrProSerGly

                TGTGAAAATATTTCTGGTGCATTTGCGAGCAATCGGACATTTGACATTACTGTCTCGGGT
1681            ----------+---------+---------+---------+---------+---------+ 1740
                CysGluAsnIleSerGlyAlaPheAlaSerAsnArgThrPheAspIleThrValSerGly

                CTTGGCACGGCCCCCAAGACACTCATTATCACACGAACGGCTACCAATGCCACCACAACA
1741            ----------+---------+---------+---------+---------+---------+ 1800
                LeuGlyThrAlaProLysThrLeuIleIleThrArgThrAlaThrAsnAlaThrThrThr

                ACCCACAAGGTTATATTCTCCAAGGCACCCGAGAGCACCACCACCTCCCCTACCTTGAAT
1801            ----------+---------+---------+---------+---------+---------+ 1860
                ThrHisLysValIlePheSerLysAlaProGluSerThrThrThrSerProThrLeuAsn

                ACAACTGGATTTGCTGATCCCAATACAACGACAGGTCTACCCAGCTCTACTCACGTGCCT
1861            ----------+---------+---------+---------+---------+---------+ 1920
                ThrThrGlyPheAlaAspProAsnThrThrThrGlyLeuProSerSerThrHisValPro

                ACCAACCTCACCGCACCTGCAAGCACAGGCCCCACTGTATCCACCGCGGATGTCACCAGC
1921            ----------+---------+---------+---------+---------+---------+ 1980
                ThrAsnLeuThrAlaProAlaSerThrGlyProThrValSerThrAlaAspValThrSer

                CCAACACCAGCCGGCACAACGTCAGGCGCATCACCGGTGACACCAAGTCCATCTCCATGG
```


1981
ProThrProAlaGlyThrThrSerGlyAlaSerProValThrProSerProSerProTrp
2040
GACAACGGCACAGAAAGTAAGGCCCCCGACATGACCAGCTCCACCTCACCAGTGACTACC
2041
AspAsnGlyThrGluSerLysAlaProAspMetThrSerSerThrSerProValThrThr
2100
CCAACCCCAAATGCCACCAGCCCCACCCCAGCAGTGACTACCCCAACCCCAAATGCCACC
2101
ProThrProAsnAlaThrSerProThrProAlaValThrThrProThrProAsnAlaThr
2160
AGCCCCACCCCAGCAGTGACTACCCCAACCCCAAATGCCACCAGCCCCACCTTGGGAAAA
2161
SerProThrProAlaValThrThrProThrProAsnAlaThrSerProThrLeuGlyLys
2220
ACAAGTCCTACCTCAGCAGTGACTACCCCAACCCCAAATGCCACCAGCCCCACCTTGGGA
2221
ThrSerProThrSerAlaValThrThrProThrProAsnAlaThrSerProThrLeuGly
2280
AAAACAAGCCCCACCTCAGCAGTGACTACCCCAACCCCAAATGCCACCAGCCCCACCTTG
2281
LysThrSerProThrSerAlaValThrThrProThrProAsnAlaThrSerProThrLeu
2340
GGAAAAACAAGCCCCACCTCAGCAGTGACTACCCCAACCCCAAATGCCACCGGCCCTACT
2341
GlyLysThrSerProThrSerAlaValThrThrProThrProAsnAlaThrGlyProThr
2400
GTGGGAGAAACAAGTCCACAGGCAAATGCCACCAACCACACCTTAGGAGGAACAAGTCCC
2401
ValGlyGluThrSerProGlnAlaAsnAlaThrAsnHisThrLeuGlyGlyThrSerPro
2460
ACCCCAGTAGTTACCAGCCAACCAAAAAATGCAACCAGTGCTGTTACCACAGGCCAACAT
2461
ThrProValValThrSerGlnProLysAsnAlaThrSerAlaValThrThrGlyGlnHis
2520
AACATAACTTCAAGTTCAACCTCTTCCATGTCACTGAGACCCAGTTCAAACCCAGAGACA
2521
AsnIleThrSerSerSerThrSerSerMetSerLeuArgProSerSerAsnProGluThr
2580
CTCAGCCCCTCCACCAGTGACAATTCAACGTCACATATGCCTTTACTAACCTCCGCTCAC
2581
LeuSerProSerThrSerAspAsnSerThrSerHisMetProLeuLeuThrSerAlaHis
2640
CCAACAGGTGGTGAAAATATAACACAGGTGACACCAGCCTCTATCAGCACACATCATGTG
2641
ProThrGlyGlyGluAsnIleThrGlnValThrProAlaSerIleSerThrHisHisVal
2700
TCCACCAGTTCGCCAGCACCCCGCCCAGGCACCACCAGCCAAGCGTCAGGCCCTGGAAAC
2701
SerThrSerSerProAlaProArgProGlyThrThrSerGlnAlaSerGlyProGlyAsn
2760
AGTTCCACATCCACAAAACCGGGGGAGGTTAATGTCACCAAAGGCACGCCCCCCCAAAAT
2761
SerSerThrSerThrLysProGlyGluValAsnValThrLysGlyThrProProGlnAsn
2820
GCAACGTCGCCCCAGGCCCCCAGTGGCCAAAAGACGGCGGTTCCCACGGTCACCTCAACA
2821
AlaThrSerProGlnAlaProSerGlyGlnLysThrAlaValProThrValThrSerThr
2880
GGTGGAAAGGCCAATTCTACCACCGGTGGAAAGCACACCACAGGACATGGAGCCCGGACA
2881
GlyGlyLysAlaAsnSerThrThrGlyGlyLysHisThrThrGlyHisGlyAlaArgThr
2940

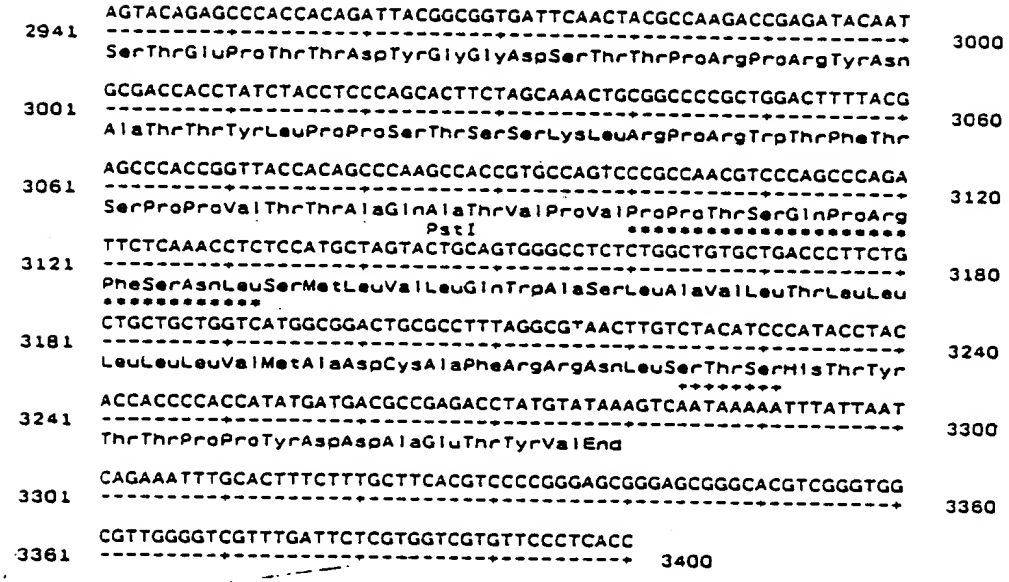
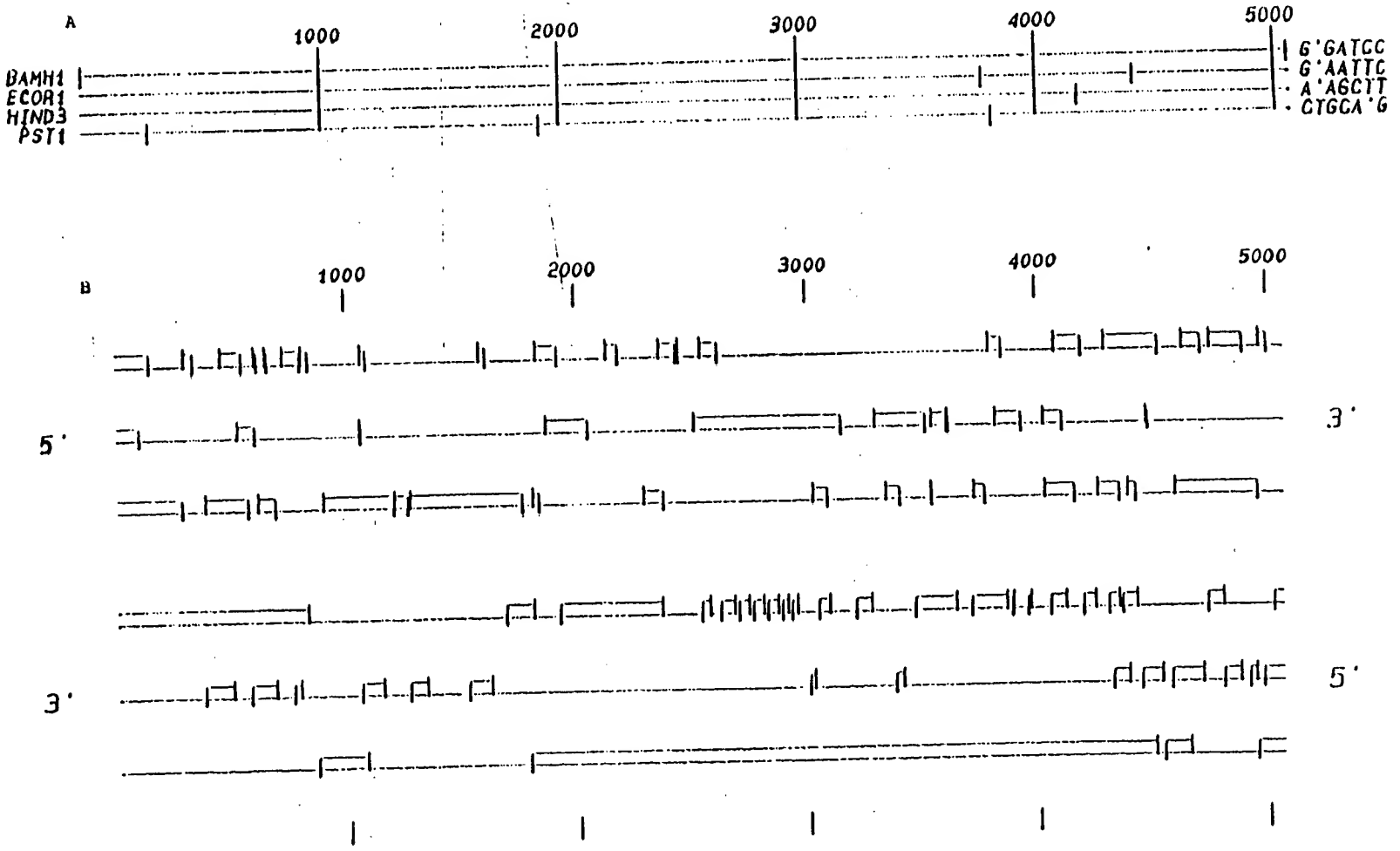

AGTACAGAGCCCACCACAGATTACGGCGGTGATTCAACTACGCCAAGACCGAGATACAAT
2941
3000
SerThrGluProThrThrAspTyrGlyGlyAspSerThrThrProArgProArgTyrAsn
GCGACCACCTATCTACCTCCCAGCACTTCTAGCAAACTGCGGCCCCGCTGGACTTTTACG
3001
3060
AlaThrThrTyrLeuProProSerThrSerSerLysLeuArgProArgTrpThrPheThr
AGCCCACCGGTTACCACAGCCCAAGCCACCGTGCCAGTCCCGCCAACGTCCCAGCCCAGA
3061
3120
SerProProValThrThrAlaGlnAlaThrValProValProProThrSerGlnProArg
PstI
TTCTCAAACCTCTCCATGCTAGTACTGCAGTGGGCCTCTCTGGCTGTGCTGACCCTTCTG
3121
3180
PheSerAsnLeuSerMetLeuValLeuGlnTrpAlaSerLeuAlaValLeuThrLeuLeu
CTGCTGCTGGTCATGGCGGACTGCGCCTTTAGGCGTAACTTGTCTACATCCCATACCTAC
3181
3240
LeuLeuLeuValMetAlaAspCysAlaPheArgArgAsnLeuSerThrSerHisThrTyr
ACCACCCCACCATATGATGACGCCGAGACCTATGTATAAAGTCAATAAAAATTTATTAAT
3241
3300
ThrThrProProTyrAspAspAlaGluThrTyrValEnd
CAGAAATTTGCACTTTCTTTGCTTCACGTCCCCGGGAGCGGGAGCGGGCACGTCGGGTGG
3301
3360
CGTTGGGGTCGTTTGATTCTCGTGGTCGTGTTCCCTCACC
3361
3400

Figure 18



0173254

Figure 19

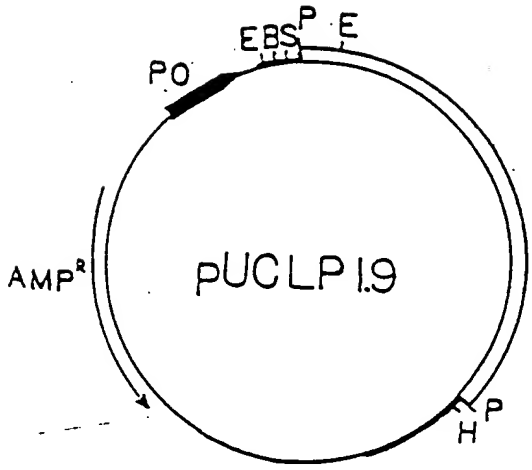


Figure 20

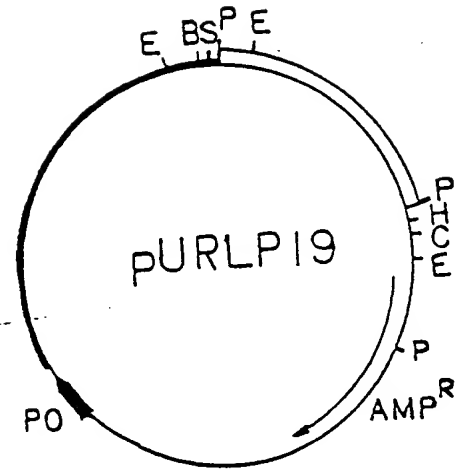


Figure 21


```
1    ACCATGATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGC  60
        MetIleThrAspSerLeuAlaValValLeuGlnArgArgAspTrpGluAsnProGly

61   GTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAA  120
     ValThrGlnLeuAsnArgLeuAlaAlaHisProProPheAlaSerTrpArgAsnSerGlu

121  GAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTT  180
     GluAlaArgThrAspArgProSerGlnGlnLeuArgSerLeuAsnGlyGluTrpArgPhe

181  GCCTGGTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAG  240
     AlaTrpPheProAlaProGluAlaValProGluSerTrpLeuGluCysAspLeuProGlu

241  GCCGATACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTAC  300
     AlaAspThrValValValProSerAsnTrpGlnMetHisGlyTyrAspAlaProIleTyr

301  ACCAACGTAACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACGGAGAATCCGACG  360
     ThrAsnValThrTyrProIleThrValAsnProProPheValProThrGluAsnProThr

361  GGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTACAGGAAGGCCAGACGCGA  420
     GlyCysTyrSerLeuThrPheAsnValAspGluSerTrpLeuGlnGluGlyGlnThrArg

421  ATTATTTTTGATGGCGTTAACTCGGCGTTTCATCTGTGGTGCAACGGGCGCTGGGTCGGT  480
     IleIlePheAspGlyValAsnSerAlaPheHisLeuTrpCysAsnGlyArgTrpValGly

481  TACGGCCAGGACAGTCGTTTGCCGTCTGAATTTGACCTGAGCGCATTTTTACGCGCCGGA  540
     TyrGlyGlnAspSerArgLeuProSerGluPheAspLeuSerAlaPheLeuArgAlaGly

541  GAAAACCGCCTCGCGGTGATGGTGCTGCGTTGGAGTGACGGCAGTTATCTGGAAGATCAG  600
     GluAsnArgLeuAlaValMetValLeuArgTrpSerAspGlySerTyrLeuGluAspGln

601  GATATGTGGCGGATGAGCGGCATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACTACA  660
     AspMetTrpArgMetSerGlyIlePheArgAspValSerLeuLeuHisLysProThrThr

661  CAAATCAGCGATTTCCATGTTGCCACTCGCTTTAATGATGATTTCAGCCGCGCTGTACTG  720
     GlnIleSerAspPheHisValAlaThrArgPheAsnAspAspPheSerArgAlaValLeu

721  GAGGCTGAAGTTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTA  780
     GluAlaGluValGlnMetCysGlyGluLeuArgAspTyrLeuArgValThrValSerLeu

781  TGGCAGGGTGAAACGCAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGAT  840
     TrpGlnGlyGluThrGlnValAlaSerGlyThrAlaProPheGlyGlyGluIleIleAsp

841  GAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCCGAAACTG  900
     GluArgGlyGlyTyrAlaAspArgValThrLeuArgLeuAsnValGluAsnProLysLeu

     TGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCACACCGCCGACGGC
```

```
901  ----------+----------+----------+----------+----------+----------+ 960
     TrpSerAlaGluIleProAsnLeuTyrArgAlaValValGluLeuHisThrAlaAspGly

     ACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGAGGTGCGGATTGAAAATGGT
961  ----------+----------+----------+----------+----------+----------+ 1020
     ThrLeuIleGluAlaGluAlaCysAspValGlyPheArgGluValArgIleGluAsnGly

     CTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCGAGGCGTTAACCGTCACGAGCATCAT
1021 ----------+----------+----------+----------+----------+----------+ 1080
     LeuLeuLeuLeuAsnGlyLysProLeuLeuIleArgGlyValAsnArgHisGluHisHis

     CCTCTGCATGGTCAGGTCATGGATGAGCAGACGATGGTGCAGGATATCCTGCTGATGAAG
1081 ----------+----------+----------+----------+----------+----------+ 1140
     ProLeuHisGlyGlnValMetAspGluGlnThrMetValGlnAspIleLeuLeuMetLys

     CAGAACAACTTTAACGCCGTGCGCTGTTCGCATTATCCGAACCATCCGCTGTGGTACACG
1141 ----------+----------+----------+----------+----------+----------+ 1200
     GlnAsnAsnPheAsnAlaValArgCysSerHisTyrProAsnHisProLeuTrpTyrThr

     CTGTGCGACCGCTACGGCCTGTATGTGGTGGATGAAGCCAATATTGAAACCCACGGCATG
1201 ----------+----------+----------+----------+----------+----------+ 1260
     LeuCysAspArgTyrGlyLeuTyrValValAspGluAlaAsnIleGluThrHisGlyMet

     GTGCCAATGAATCGTCTGACCGATGATCCGCGCTGGCTACCGGCGATGAGCGAACGCGTA
1261 ----------+----------+----------+----------+----------+----------+ 1320
     ValProMetAsnArgLeuThrAspAspProArgTrpLeuProAlaMetSerGluArgVal

     ACGCGAATGGTGCAGCGCGATCGTAATCACCCGAGTGTGATCATCTGGTCGCTGGGGAAT
1321 ----------+----------+----------+----------+----------+----------+ 1380
     ThrArgMetValGlnArgAspArgAsnHisProSerValIleIleTrpSerLeuGlyAsn

     GAATCAGGCCACGGCGCTAATCACGACGCGCTGTATCGCTGGATCAAATCTGTCGATCCT
1381 ----------+----------+----------+----------+----------+----------+ 1440
     GluSerGlyHisGlyAlaAsnHisAspAlaLeuTyrArgTrpIleLysSerValAspPro

     TCCCGCCCGGTGCAGTATGAAGGCGGCGGAGCCGACACCACGGCCACCGATATTATTTGC
1441 ----------+----------+----------+----------+----------+----------+ 1500
     SerArgProValGlnTyrGluGlyGlyGlyAlaAspThrThrAlaThrAspIleIleCys

     CCGATGTACGCGCGCGTGGATGAAGACCAGCCCTTCCCGGCTGTGCCGAAATGGTCCATC
1501 ----------+----------+----------+----------+----------+----------+ 1560
     ProMetTyrAlaArgValAspGluAspGlnProPheProAlaValProLysTrpSerIle

     AAAAAATGGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCAC
1561 ----------+----------+----------+----------+----------+----------+ 1620
     LysLysTrpLeuSerLeuProGlyGluThrArgProLeuIleLeuCysGluTyrAlaHis

     GCGATGGGTAACAGTCTTGGCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCC
1621 ----------+----------+----------+----------+----------+----------+ 1680
     AlaMetGlyAsnSerLeuGlyGlyPheAlaLysTyrTrpGlnAlaPheArgGlnTyrPro

     CGTTTACAGGGCGGCTTCGTCTGGGACTGGGTGGATCAGTCGCTGATTAAATATGATGAA
1681 ----------+----------+----------+----------+----------+----------+ 1740
     ArgLeuGlnGlyGlyPheValTrpAspTrpValAspGlnSerLeuIleLysTyrAspGlu

     AACGGCAACCCGTGGTCGGCTTACGGCGGTGATTTTGGCGATACGCCGAACGATCGCCAG
1741 ----------+----------+----------+----------+----------+----------+ 1800
     AsnGlyAsnProTrpSerAlaTyrGlyGlyAspPheGlyAspThrProAsnAspArgGln

     TTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAGCGCTGACGGAAGCA
1801 ----------+----------+----------+----------+----------+----------+ 1860
     PheCysMetAsnGlyLeuValPheAlaAspArgThrProHisProAlaLeuThrGluAla
```

```
                AAACACCAGCAGCAGTTTTTTCCAGTTCCGTTTATCCGGGCAAACCATCGAAGTGACCAGC
1861            ---------+---------+---------+---------+---------+---------+ 1920
                LysHisGlnGlnGlnPhePheGlnPheArgLeuSerGlyGlnThrIleGluValThrSer

                GAATACCTGTTCCGTCATAGCGATAACGAGCTCCTGCACTGGATGGTGGCGCTGGATGGT
1921            ---------+---------+---------+---------+---------+---------+ 1980
                GluTyrLeuPheArgHisSerAspAsnGluLeuLeuHisTrpMetValAlaLeuAspGly

                AAGCCGCTGGCAAGCGGTGAAGTGCCTCTGGATGTCGCTCCACAAGGTAAACAGTTGATT
1981            ---------+---------+---------+---------+---------+---------+ 2040
                LysProLeuAlaSerGlyGluValProLeuAspValAlaProGlnGlyLysGlnLeuIle

                GAACTGCCTGAACTACCGCAGCCGGAGAGCGCCGGGCAACTCTGGCTCACAGTACGCGTA
2041            ---------+---------+---------+---------+---------+---------+ 2100
                GluLeuProGluLeuProGlnProGluSerAlaGlyGlnLeuTrpLeuThrValArgVal

                GTGCAACCGAACGCGACCGCATGGTCAGAAGCCGGGCACATCAGCGCCTGGCAGCAGTGG
2101            ---------+---------+---------+---------+---------+---------+ 2160
                ValGlnProAsnAlaThrAlaTrpSerGluAlaGlyHisIleSerAlaTrpGlnGlnTrp

                CGTCTGGCGGAAAACCTCAGTGTGACGCTCCCCGCCGCGTCCCACGCCATCCCGCATCTG
2161            ---------+---------+---------+---------+---------+---------+ 2220
                ArgLeuAlaGluAsnLeuSerValThrLeuProAlaAlaSerHisAlaIleProHisLeu

                ACCACCAGCGAAATGGATTTTTGCATCGAGCTGGGTAATAAGCGTTGGCAATTTAACCGC
2221            ---------+---------+---------+---------+---------+---------+ 2280
                ThrThrSerGluMetAspPheCysIleGluLeuGlyAsnLysArgTrpGlnPheAsnArg

                CAGTCAGGCTTTCTTTCACAGATGTGGATTGGCGATAAAAAACAACTGCTGACGCCGCTG
2281            ---------+---------+---------+---------+---------+---------+ 2340
                GlnSerGlyPheLeuSerGlnMetTrpIleGlyAspLysLysGlnLeuLeuThrProLeu

                CGCGATCAGTTCACCCGTGCACCGCTGGATAACGACATTGGCGTAAGTGAAGCGACCCGC
2341            ---------+---------+---------+---------+---------+---------+ 2400
                ArgAspGlnPheThrArgAlaProLeuAspAsnAspIleGlyValSerGluAlaThrArg

                ATTGACCCTAACGCCTGGGTCGAACGCTGGAAGGCGGCGGGCCATTACCAGGCCGAAGCA
2401            ---------+---------+---------+---------+---------+---------+ 2460
                IleAspProAsnAlaTrpValGluArgTrpLysAlaAlaGlyHisTyrGlnAlaGluAla

                GCGTTGTTGCAGTGCACGGCAGATACACTTGCTGATGCGGTGCTGATTACGACCGCTCAC
2461            ---------+---------+---------+---------+---------+---------+ 2520
                AlaLeuLeuGlnCysThrAlaAspThrLeuAlaAspAlaValLeuIleThrThrAlaHis

                GCGTGGCAGCATCAGGGGAAAACCTTATTTATCAGCCGGAAAACCTACCGGATTGATGGT
2521            ---------+---------+---------+---------+---------+---------+ 2580
                AlaTrpGlnHisGlnGlyLysThrLeuPheIleSerArgLysThrTyrArgIleAspGly

                AGTGGTCAAATGGCGATTACCGTTGATGTTGAAGTGGCGAGCGATACACCGCATCCGGCG
2581            ---------+---------+---------+---------+---------+---------+ 2640
                SerGlyGlnMetAlaIleThrValAspValGluValAlaSerAspThrProHisProAla

                CGGATTGGCCTGAACTGCCAGCTGGCGCAGGTAGCAGAGCGGGTAAACTGGCTCGGATTA
2641            ---------+---------+---------+---------+---------+---------+ 2700
                ArgIleGlyLeuAsnCysGlnLeuAlaGlnValAlaGluArgValAsnTrpLeuGlyLeu

                GGGCCGCAAGAAAACTATCCCGACCGCCTTACTGCCGCCTGTTTTGACCGCTGGGATCTG
2701            ---------+---------+---------+---------+---------+---------+ 2760
                GlyProGlnGluAsnTyrProAspArgLeuThrAlaAlaCysPheAspArgTrpAspLeu

                CCATTGTCAGACATGTATACCCCGTACGTCTTCCCGAGCGAAAACGGTCTGCGCTGCGGG
2761            ---------+---------+---------+---------+---------+---------+ 2820
                ProLeuSerAspMetTyrThrProTyrValPheProSerGluAsnGlyLeuArgCysGly
```

```
2821 ACGCGCGAATTGAATTATGGCCCACACCAGTGGCGCGGCGACTTCCAGTTCAACATCAGC 2880
     ThrArgGluLeuAsnTyrGlyProHisGlnTrpArgGlyAspPheGlnPheAsnIleSer

2881 CGCTACAGTCAACAGCAACTGATGGAAACCAGCCATCGCCATCTGCTGCACGCGGAAGAA 2940
     ArgTyrSerGlnGlnGlnLeuMetGluThrSerHisArgHisLeuLeuHisAlaGluGlu

2941 GGCACATGGCTGAATATCGACGGTTTCCATATGGGGATTGGTGGCGACGACTCCTGGAGC 3000
     GlyThrTrpLeuAsnIleAspGlyPheHisMetGlyIleGlyGlyAspAspSerTrpSer

3001 CCGTCAGTATCGGCGGAATTCCAGCTGAGCGCCGGTCGCTACCATTACCAGTTGGTCTGG 3060
     ProSerValSerAlaGluPheGlnLeuSerAlaGlyArgTyrHisTyrGlnLeuValTrp

3061 TGTCAAAAAggggatccgtcgacctgcaGTGGATACGAGAGCCATGTTCCCAGCGGCGGA 3120
     CysGlnLysGlyAspProSerThrCysSerGlyTyrGluSerHisValProSerGlyGly

3121 ATTCTCACATCAACGAGTCCCGTGGCCACCCCAATACCTGGTACAGGGTATGCATACAGC 3180
     IleLeuThrSerThrSerProValAlaThrProIleProGlyThrGlyTyrAlaTyrSer

3181 CTGCGTCTGACACCACGTCCAGTGTCACGATTTCTTGGCAATAACAGTATCCTGTACGTG 3240
     LeuArgLeuThrProArgProValSerArgPheLeuGlyAsnAsnSerIleLeuTyrVal

3241 TTTTACTCTGGGAATGGACCGAAGGCGAGCGGGGGAGATTACTGCATTCAGTCCAACATT 3300
     PheTyrSerGlyAsnGlyProLysAlaSerGlyGlyAspTyrCysIleGlnSerAsnIle

3301 GTGTTCTCTGATGAGATTCCAGCTTCACAGGACATGCCGACAAACACCACAGACATCACA 3360
     ValPheSerAspGluIleProAlaSerGlnAspMetProThrAsnThrThrAspIleThr

3361 TATGTGGGTGACAATGCTACCTATTCAGTGCCAATGGTCACTTCTGAGGACGCAAACTCG 3420
     TyrValGlyAspAsnAlaThrTyrSerValProMetValThrSerGluAspAlaAsnSer

3421 CCAAATGTTACAGTGACTGCCTTTTGGGCCTGGCCAAACAACACTGAAACTGACTTTAAG 3480
     ProAsnValThrValThrAlaPheTrpAlaTrpProAsnAsnThrGluThrAspPheLys

3481 TGCAAATGGACTCTCACCTCGGGGACACCTTCGGGTTGTGAAAATATTTCTGGTGCATTT 3540
     CysLysTrpThrLeuThrSerGlyThrProSerGlyCysGluAsnIleSerGlyAlaPhe

3541 GCGAGCAATCGGACATTTGACATTACTGTCTCGGGTCTTGGCACGGCCCCCAAGACACTC 3600
     AlaSerAsnArgThrPheAspIleThrValSerGlyLeuGlyThrAlaProLysThrLeu

3601 ATTATCACACGAACGGCTACCAATGCCACCACAACAACCCACAAGGTTATATTCTCCAAG 3660
     IleIleThrArgThrAlaThrAsnAlaThrThrThrThrHisLysValIlePheSerLys

3661 GCACCCGAGAGCACCACCACCTCCCCTACCTTGAATACAACTGGATTTGCTGATCCCAAT 3720
     AlaProGluSerThrThrThrSerProThrLeuAsnThrThrGlyPheAlaAspProAsn

3721 ACAACGACAGGTCTACCCAGCTCTACTCACGTGCCTACCAACCTCACCGCACCTGCAAGC 3780
```

```
                ThrThrThrGlyLeuProSerSerThrHisValProThrAsnLeuThrAlaProAlaSer

                ACAGGCCCCACTGTATCCACCGCGGATGTCACCAGCCCAACACCAGCCGGCACAACGTCA
3781            ----------+---------+---------+---------+---------+---------- 3840
                ThrGlyProThrValSerThrAlaAspValThrSerProThrProAlaGlyThrThrSer

                GGCGCATCACCGGTGACACCAAGTCCATCTCCATGGGACAACGGCACAGAAAGTAAGGCC
3841            ----------+---------+---------+---------+---------+---------- 3900
                GlyAlaSerProValThrProSerProSerProTrpAspAsnGlyThrGluSerLysAla

                CCCGACATGACCAGCTCCACCTCACCAGTGACTACCCCAACCCCAAATGCCACCAGCCCC
3901            ----------+---------+---------+---------+---------+---------- 3960
                ProAspMetThrSerSerThrSerProValThrThrProThrProAsnAlaThrSerPro

                ACCCCAGCAGTGACTACCCCAACCCCAAATGCCACCAGCCCCACCCCAGCAGTGACTACC
3961            ----------+---------+---------+---------+---------+---------- 4020
                ThrProAlaValThrThrProThrProAsnAlaThrSerProThrProAlaValThrThr

                CCAACCCCAAATGCCACCAGCCCCACCTTGGGAAAAACAAGTCCTACCTCAGCAGTGACT
4021            ----------+---------+---------+---------+---------+---------- 4080
                ProThrProAsnAlaThrSerProThrLeuGlyLysThrSerProThrSerAlaValThr

                ACCCCAACCCCAAATGCCACCAGCCCCACCTTGGGAAAAACAAGCCCCACCTCAGCAGTG
4081            ----------+---------+---------+---------+---------+---------- 4140
                ThrProThrProAsnAlaThrSerProThrLeuGlyLysThrSerProThrSerAlaVal

                ACTACCCCAACCCCAAATGCCACCAGCCCCACCTTGGGAAAAACAAGCCCCACCTCAGCA
4141            ----------+---------+---------+---------+---------+---------- 4200
                ThrThrProThrProAsnAlaThrSerProThrLeuGlyLysThrSerProThrSerAla

                GTGACTACCCCAACCCCAAATGCCACCGGCCCTACTGTGGGAGAAACAAGTCCACAGGCA
4201            ----------+---------+---------+---------+---------+---------- 4260
                ValThrThrProThrProAsnAlaThrGlyProThrValGlyGluThrSerProGlnAla

                AATGCCACCAACCACACCTTAGGAGGAACAAGTCCCACCCCAGTAGTTACCAGCCAACCA
4261            ----------+---------+---------+---------+---------+---------- 4320
                AsnAlaThrAsnHisThrLeuGlyGlyThrSerProThrProValValThrSerGlnPro

                AAAAATGCAACCAGTGCTGTTACCACAGGCCAACATAACATAACTTCAAGTTCAACCTCT
4321            ----------+---------+---------+---------+---------+---------- 4380
                LysAsnAlaThrSerAlaValThrThrGlyGlnHisAsnIleThrSerSerSerThrSer

                TCCATGTCACTGAGACCCAGTTCAAACCCAGAGACACTCAGCCCCTCCACCAGTGACAAT
4381            ----------+---------+---------+---------+---------+---------- 4440
                SerMetSerLeuArgProSerSerAsnProGluThrLeuSerProSerThrSerAspAsn

                TCAACGTCACATATGCCTTTACTAACCTCCGCTCACCCAACAGGTGGTGAAAATATAACA
4441            ----------+---------+---------+---------+---------+---------- 4500
                SerThrSerHisMetProLeuLeuThrSerAlaHisProThrGlyGlyGluAsnIleThr

                CAGGTGACACCAGCCTCTATCAGCACACATCATGTGTCCACCAGTTCGCCAGCACCCCGC
4501            ----------+---------+---------+---------+---------+---------- 4560
                GlnValThrProAlaSerIleSerThrHisHisValSerThrSerSerProAlaProArg

                CCAGGCACCACCAGCCAAGCGTCAGGCCCTGGAAACAGTTCCACATCCACAAAACCGGGG
4561            ----------+---------+---------+---------+---------+---------- 4620
                ProGlyThrThrSerGlnAlaSerGlyProGlyAsnSerSerThrSerThrLysProGly

                GAGGTTAATGTCACCAAAGGCACGCCCCCCCAAAATGCAACGTCGCCCCAGGCCCCCAGT
4621            ----------+---------+---------+---------+---------+---------- 4680
                GluValAsnValThrLysGlyThrProProGlnAsnAlaThrSerProGlnAlaProSer

                GGCCAAAAGACGGCGGTTCCCACGGTCACCTCAACAGGTGGAAAGGCCAATTCTACCACC
```

```
4681  ---------+---------+---------+---------+---------+---------+  4740
      GlyGlnLysThrAlaValProThrValThrSerThrGlyGlyLysAlaAsnSerThrThr

      GGTGGAAAGCACACCACAGGACATGGAGCCCGGACAAGTACAGAGCCCACCACAGATTAC
4741  ---------+---------+---------+---------+---------+---------+  4800
      GlyGlyLysHisThrThrGlyHisGlyAlaArgThrSerThrGluProThrThrAspTyr

      GGCGGTGATTCAACTACGCCAAGACCGAGATACAATGCGACCACCTATCTACCTCCCAGC
4801  ---------+---------+---------+---------+---------+---------+  4860
      GlyGlyAspSerThrThrProArgProArgTyrAsnAlaThrThrTyrLeuProProSer

      ACTTCTAGCAAACTGCGGCCCCGCTGGACTTTTACGAGCCCACCGGTTACCACAGCCCAA
4861  ---------+---------+---------+---------+---------+---------+  4920
      ThrSerSerLysLeuArgProArgTrpThrPheThrSerProProValThrThrAlaGln

      GCCACCGTGCCAGTCCCGCCAACGTCCCAGCCCAGATTCTCAAACCTCTCCATGCTAGTA
4921  ---------+---------+---------+---------+---------+---------+  4980
      AlaThrValProValProProThrSerGlnProArgPheSerAsnLeuSerMetLeuVal

      CTGCAGccaagcttATCGATGATAAGCTGTCAAACATGA
4981  ---------+---------+---------+--------  5019
      LeuGlnProSerLeuSerMetIleSerCysGlnThrEnd
```

Figure 22

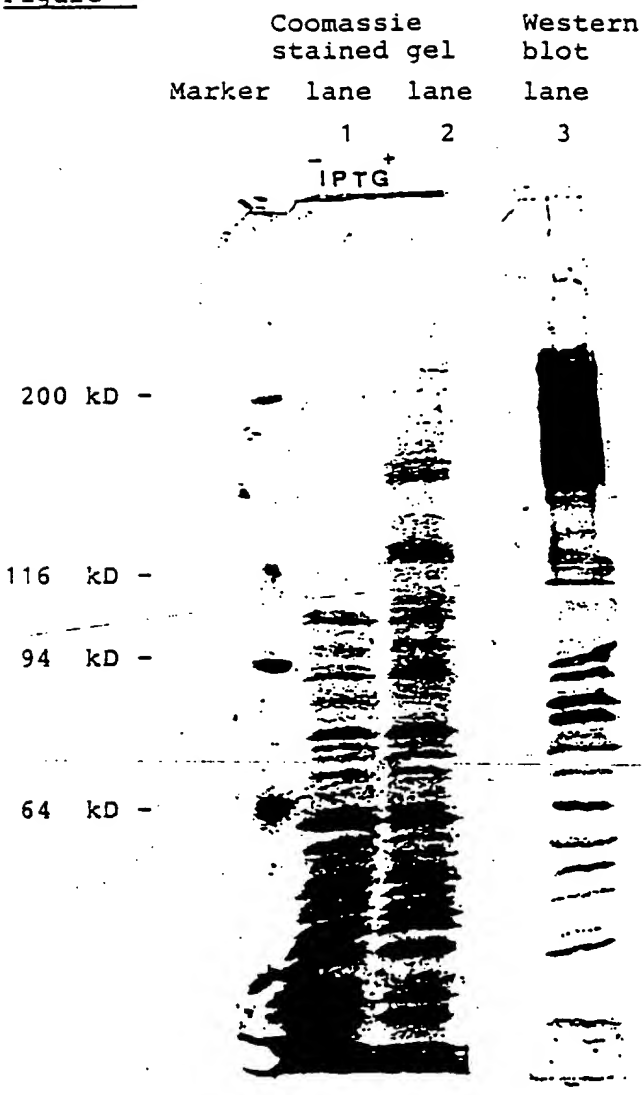

Coomassie stained gel
Western blot
Marker lane
lane 1
lane 2
lane 3
- IPTG +
200 kD -
116 kD -
94 kD -
64 kD -

Figure 23

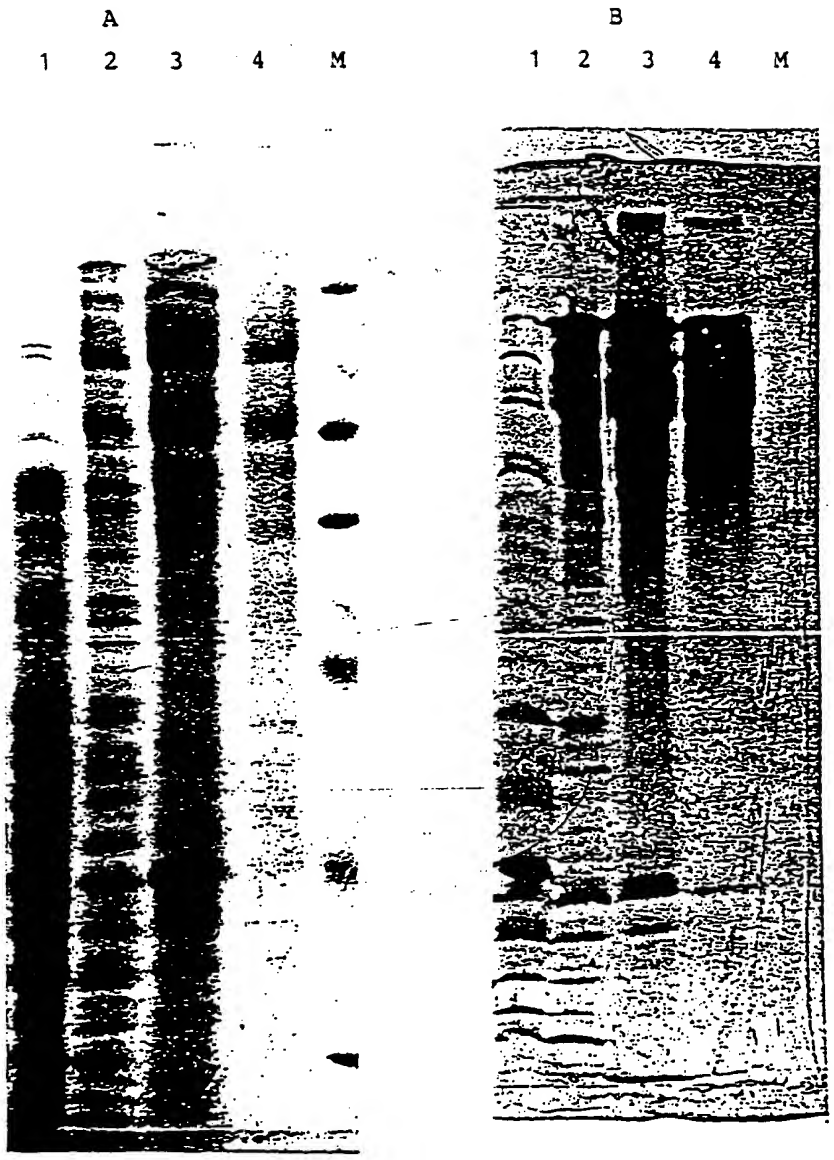


Figure 24

Figure 25:

Expression of gp350-fragments as β-gal fusion proteins

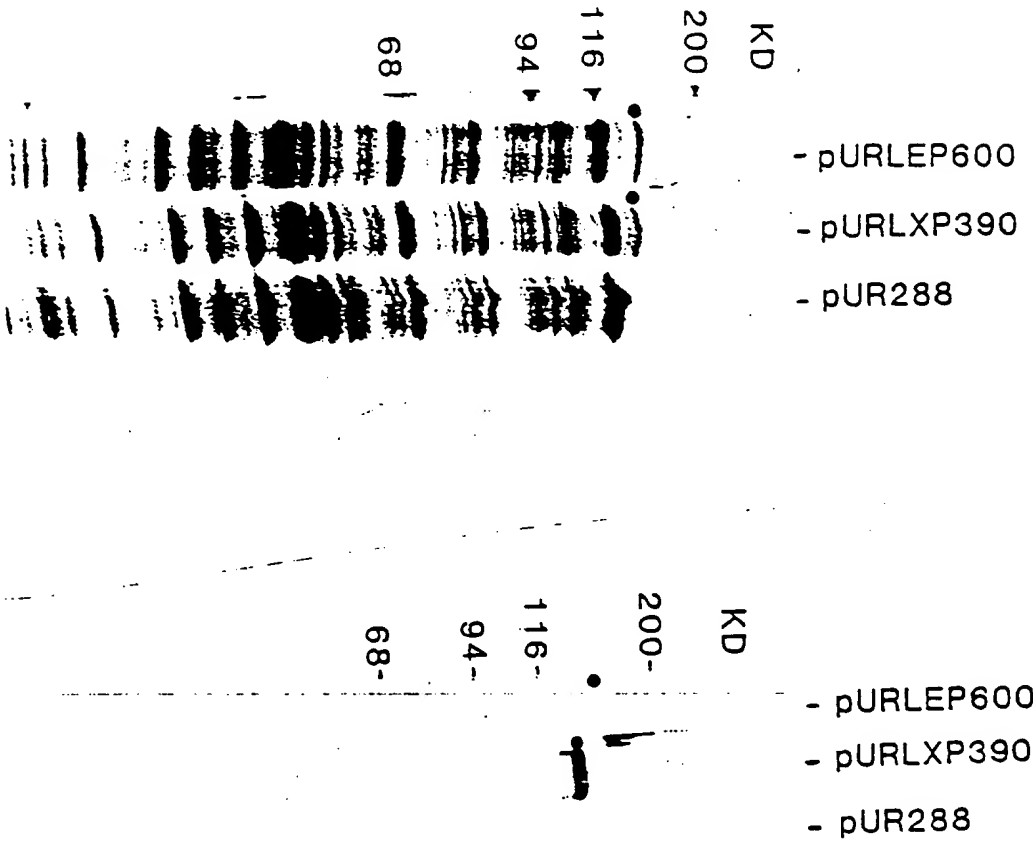
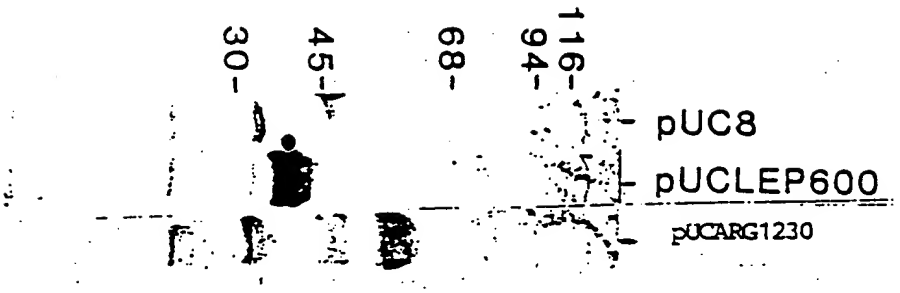


Figure 26:

Expression of proteins



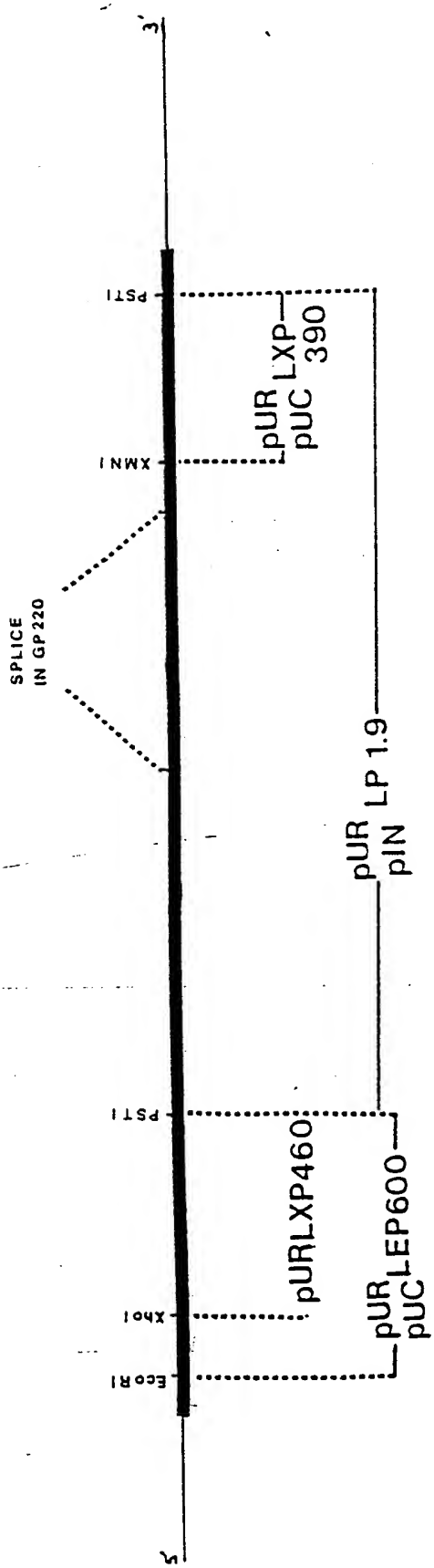
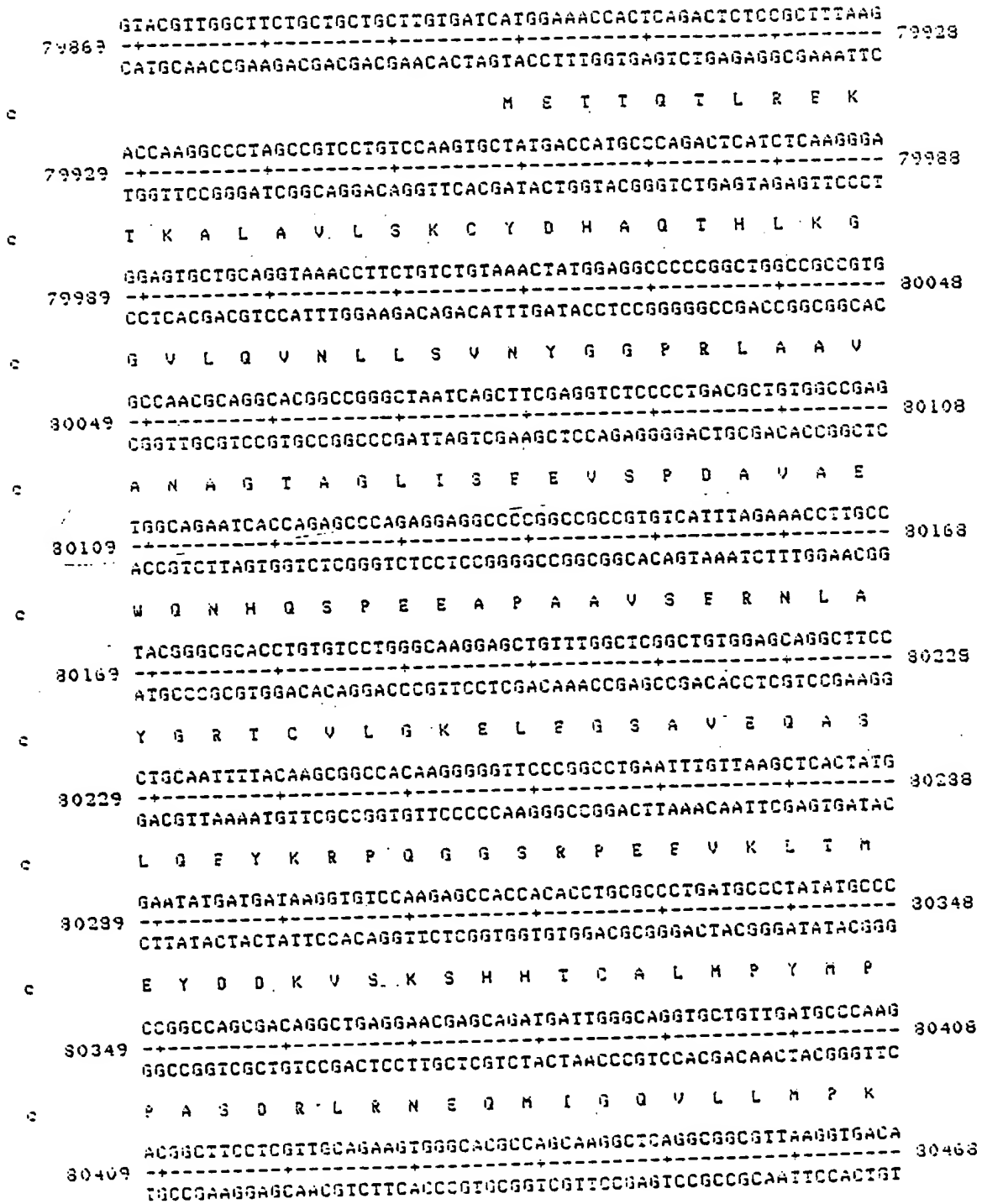
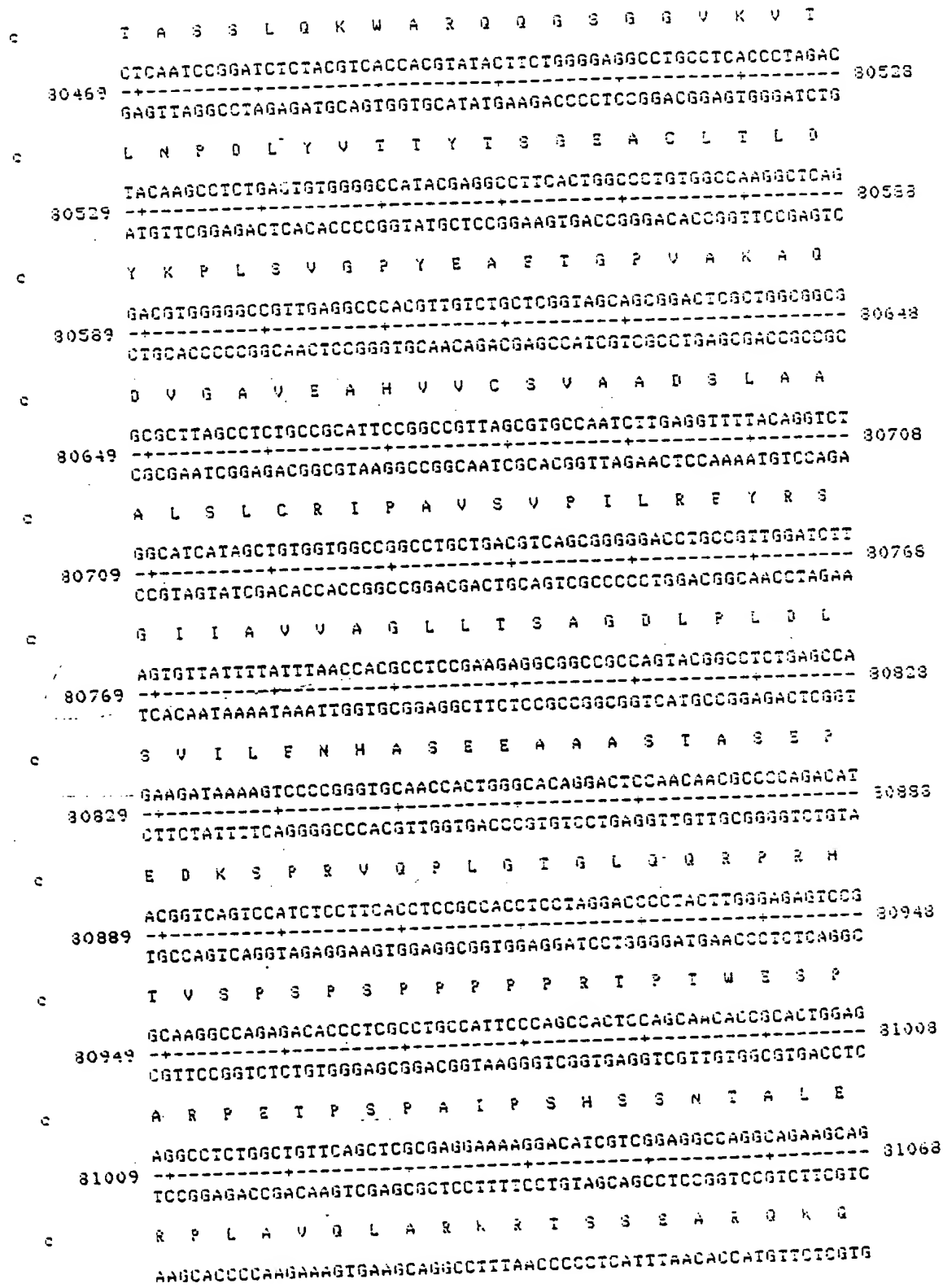


Figure 27: Restriction map of the coding region of gp 250/350

Figure 28

A: p47

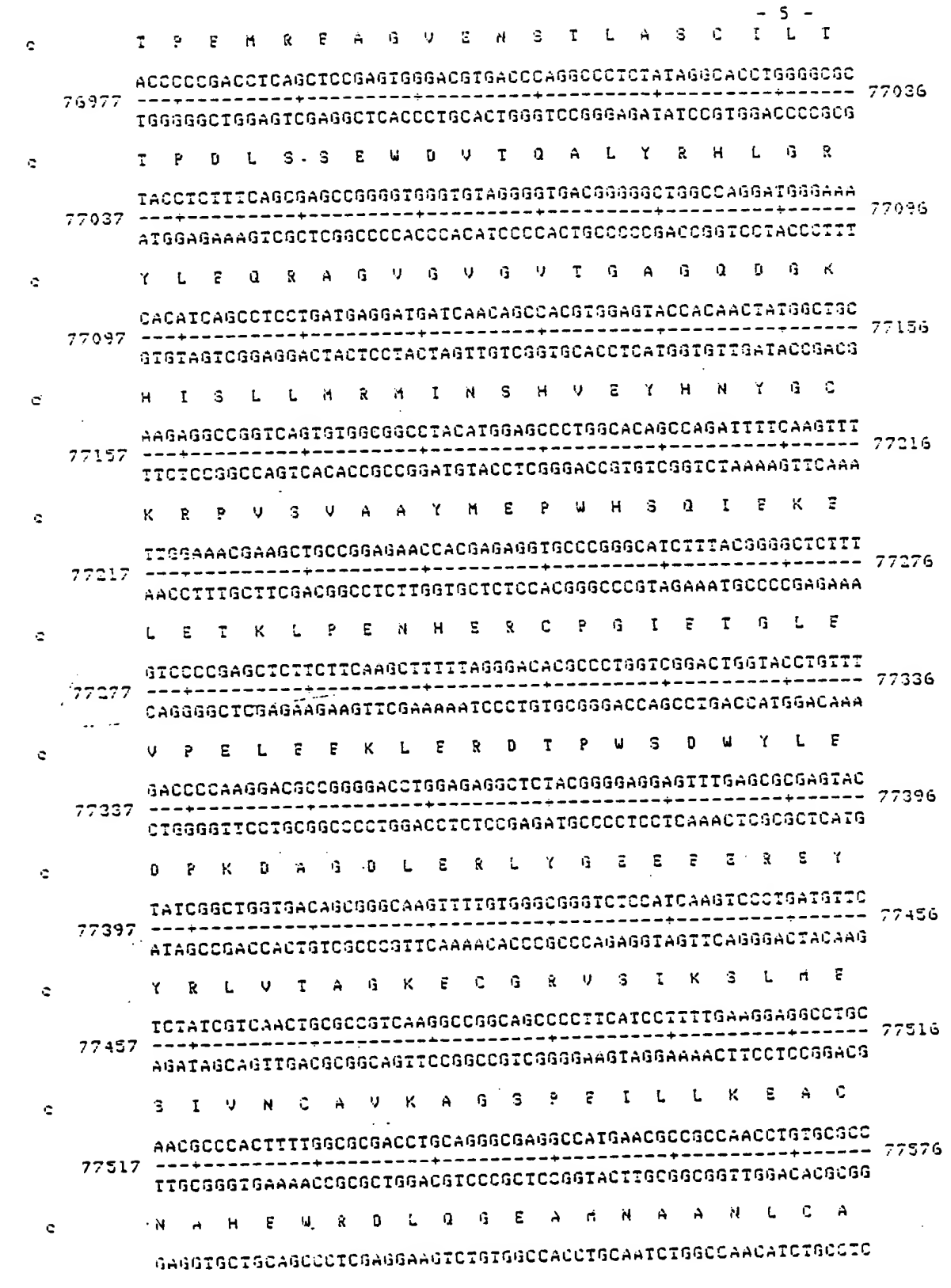


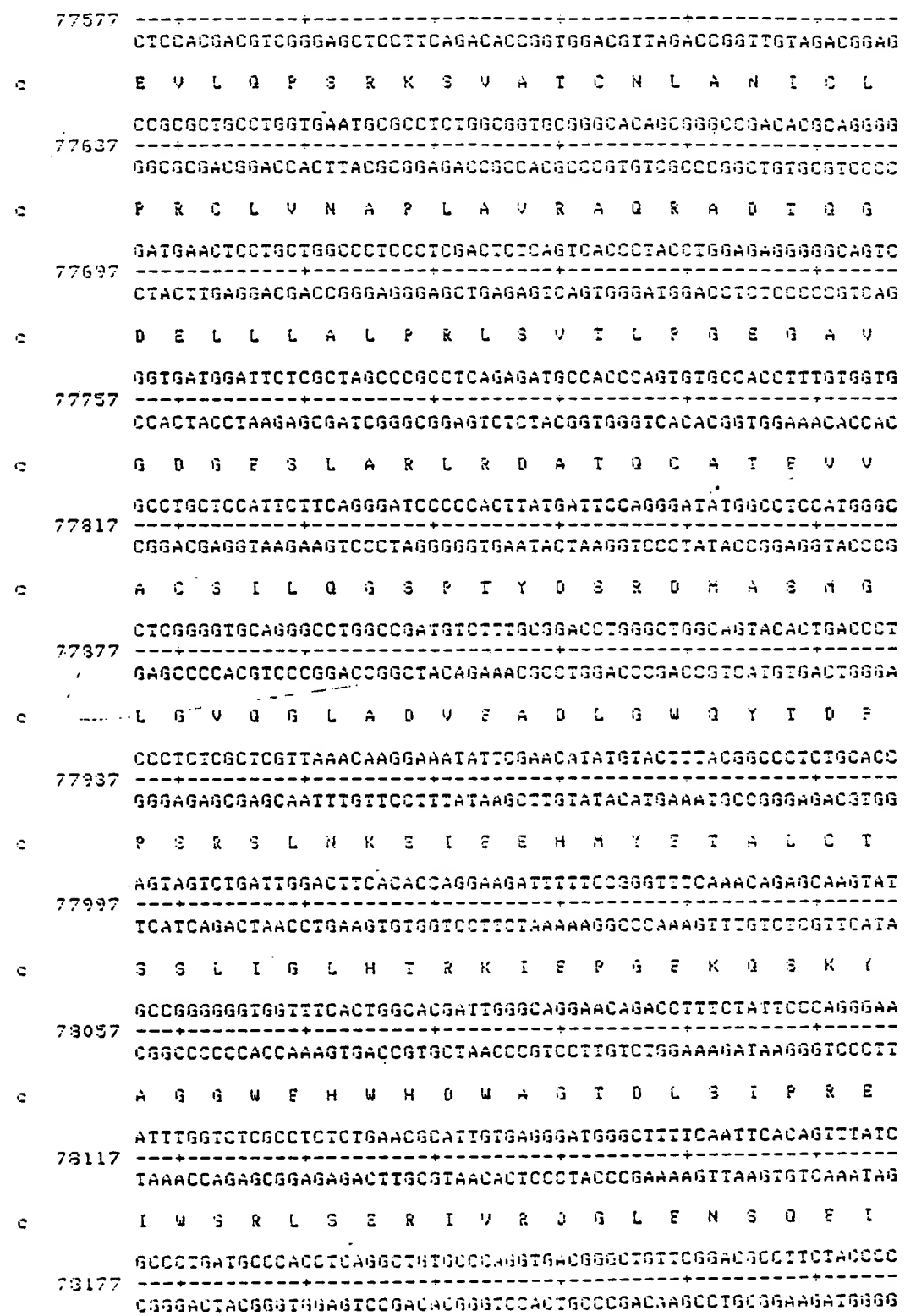

I A S S L Q K W A R Q Q G S G G V K V I
CTCAATCCGGATCTCTACGTCACCACGTATACTTCTGGGGAGGCCTGCCTCACCCTAGAC
30469 -+---------+---------+---------+---------+---------+---------+-------- 30528
GAGTTAGGCCTAGAGATGCAGTGGTGCATATGAAGACCCCTCCGGACGGAGTGGGATCTG
L N P D L Y V T I Y I S G E A C L I L D
TACAAGCCTCTGAGTGTGGGGCCATACGAGGCCTTCACTGGCCCTGTGGCCAAGGCTCAG
30529 -+---------+---------+---------+---------+---------+---------+-------- 30588
ATGTTCGGAGACTCACACCCCGGTATGCTCCGGAAGTGACCGGGACACCGGTTCCGAGTC
Y K P L S V G P Y E A E T G P V A K A Q
GACGTGGGGGCCGTTGAGGCCCACGTTGTCTGCTCGGTAGCAGCGGACTCGCTGGCGGCG
30589 -+---------+---------+---------+---------+---------+---------+-------- 30648
CTGCACCCCCGGCAACTCCGGGTGCAACAGACGAGCCATCGTCGCCTGAGCGACCGCCGC
D V G A V E A H V V C S V A A D S L A A
GCGCTTAGCCTCTGCCGCATTCCGGCCGTTAGCGTGCCAATCTTGAGGTTTTACAGGTCT
30649 -+---------+---------+---------+---------+---------+---------+-------- 30708
CGCGAATCGGAGACGGCGTAAGGCCGGCAATCGCACGGTTAGAACTCCAAAATGTCCAGA
A L S L C R I P A V S V P I L R E Y R S
GGCATCATAGCTGTGGTGGCCGGCCTGCTGACGTCAGCGGGGGACCTGCCGTTGGATCTT
30709 -+---------+---------+---------+---------+---------+---------+-------- 30768
CCGTAGTATCGACACCACCGGCCGGACGACTGCAGTCGCCCCCTGGACGGCAACCTAGAA
G I I A V V A G L L T S A G D L P L D L
AGTGTTATTTTATTTAACCACGCCTCCGAAGAGGCGGCCGCCAGTACGGCCTCTGAGCCA
30769 -+---------+---------+---------+---------+---------+---------+-------- 30828
TCACAATAAAATAAATTGGTGCGGAGGCTTCTCCGCCGGCGGTCATGCCGGAGACTCGGT
S V I L E N H A S E E A A A S I A S E P
GAAGATAAAAGTCCCCGGGTGCAACCACTGGGCACAGGACTCCAACAACGCCCCAGACAT
30829 -+---------+---------+---------+---------+---------+---------+-------- 30888
CTTCTATTTTCAGGGGCCCACGTTGGTGACCCGTGTCCTGAGGTTGTTGCGGGGTCTGTA
E D K S P R V Q P L G T G L Q Q R P R H
ACGGTCAGTCCATCTCCTTCACCTCCGCCACCTCCTAGGACCCCTACTTGGGAGAGTCCG
30889 -+---------+---------+---------+---------+---------+---------+-------- 30948
TGCCAGTCAGGTAGAGGAAGTGGAGGCGGTGGAGGATCCTGGGGATGAACCCTCTCAGGC
T V S P S P S P P P P P R I P I W E S P
GCAAGGCCAGAGACACCCTCGCCTGCCATTCCCAGCCACTCCAGCAACACCGCACTGGAG
30949 -+---------+---------+---------+---------+---------+---------+-------- 31008
CGTTCCGGTCTCTGTGGGAGCGGACGGTAAGGGTCGGTGAGGTCGTTGTGGCGTGACCTC
A R P E T P S P A I P S H S S N T A L E
AGGCCTCTGGCTGTTCAGCTCGCGAGGAAAAGGACATCGTCGGAGGCCAGGCAGAAGCAG
31009 -+---------+---------+---------+---------+---------+---------+-------- 31068
TCCGGAGACCGACAAGTCGAGCGCTCCTTTTCCTGTAGCAGCCTCCGGTCCGTCTTCGTC
R P L A V Q L A R K R T S S E A R Q K G
AAGCACCCCAAGAAAGTGAAGCAGGCCTTTAACCCCCTCATTTAACACCATGTTCTCGTG

```
   81069 -+---------+---------+---------+---------+---------+--------- 81128
         TTCGTGGGGTTCTTTCACTTCGTCCGGAAATTGGGGGAGTAAATTGTGGTACAAGAGCAC

c        K  H  P  K  K  V  K  Q  A  E  N  P  L  I  *

         CAAGCAGCACCT
   81129 -+----------+ 81140
         GTTCGTCGTGGA

c
```

B: p90

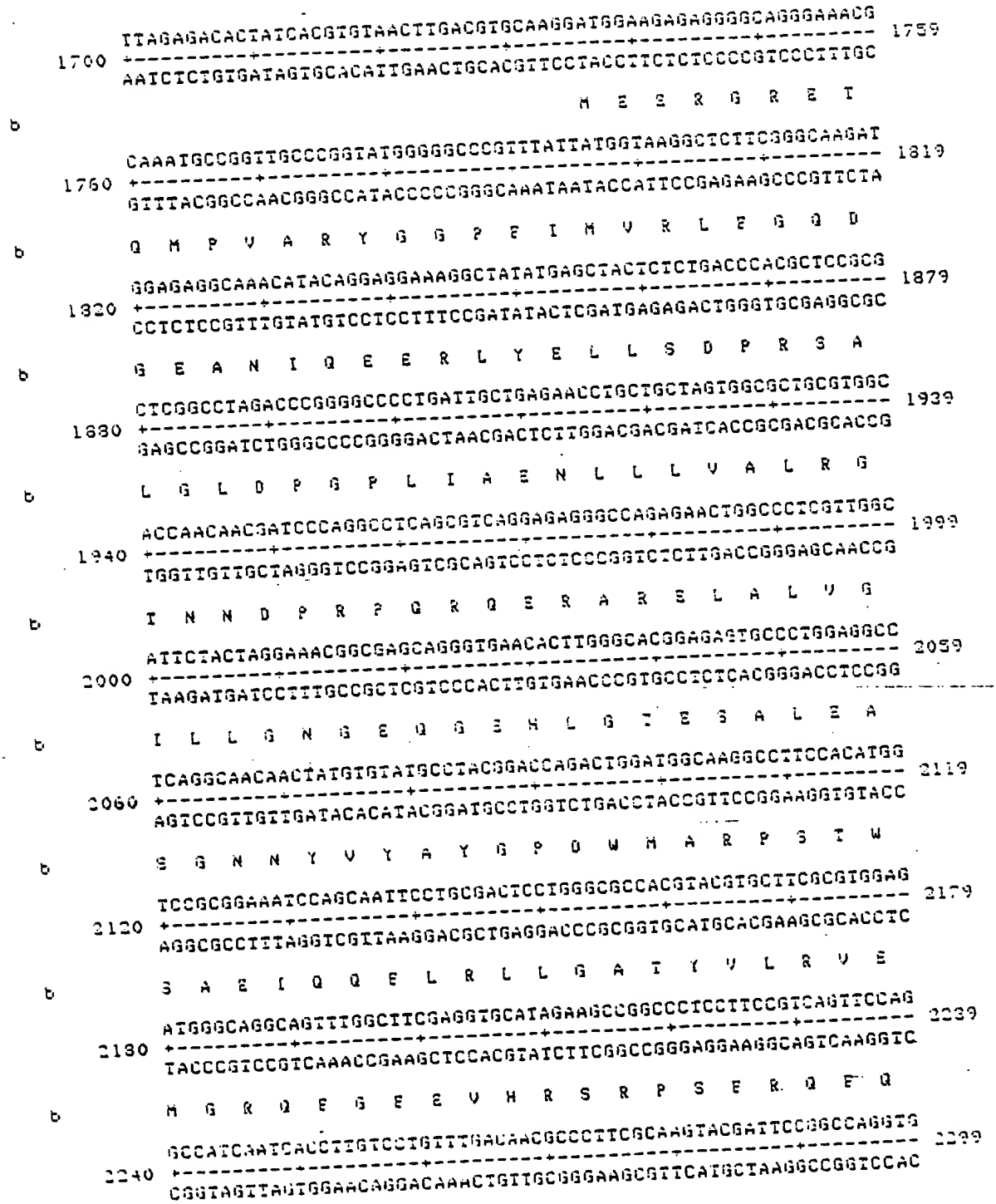

```
          AAGTGGTTCAGTGGACACCCACCACACAGCATGGCAACGACCAGTCATGTCGAGCATGAG
76377     ---+---------+---------+---------+---------+---------+---------+------ 76436
          TTCACCAAGTCACCTGTGGGTGGTGTGTCGTACCGTTGCTGGTCAGTACAGCTCGTACTC

c                                       M   A   I   T   S   H   V   E   H   E

          CTCCTCTCCAAATTGATTGATGAGTTAAAGGTCAAGGCCAACTCAGACCCCGAGGCTGAT
76437     ---+---------+---------+---------+---------+---------+---------+------ 76496
          GAGGAGAGGTTTAACTAACTACTCAATTTCCAGTTCCGGTTGAGTCTGGGGCTCCGACTA

c         L   L   S   K   L   I   D   E   L   K   V   K   A   N   S   D   P   E   A   D

          GTCCTGGCCGGGCGCCTGCTCCACCGCCTTAAGGCCGAGTCAGTTACACACACAGTAGCC
76497     ---+---------+---------+---------+---------+---------+---------+------ 76556
          CAGGACCGGCCCGCGGACGAGGTGGCGGAATTCCGGCTCAGTCAATGTGTGTGTCATCGG

c         V   L   A   G   R   L   L   H   R   L   K   A   E   S   V   T   H   T   V   A

          GAATATCTGGAGGTCTTCTCTGACAAATTCTACGATGAGGAATTCTTCCAGATGCACCGG
76557     ---+---------+---------+---------+---------+---------+---------+------ 76616
          CTTATAGACCTCCAGAAGAGACTGTTTAAGATGCTACTCCTTAAGAAGGTCTACGTGGCC

c         E   Y   L   E   V   F   S   D   K   F   Y   D   E   E   F   E   Q   M   H   R

          GATGAGCTGGAGACCCGAGTCTCTGCTTTCGCGCAGAGCCCGGCCTACGAGCGCATCGTC
76617     ---+---------+---------+---------+---------+---------+---------+------ 76676
          CTACTCGACCTCTGGGCTCAGAGACGAAAGCGCGTCTCGGGCCGGATGCTCGCGTAGCAG

c         D   E   L   E   T   R   V   S   A   F   A   Q   S   P   A   Y   E   R   I   V

          TCCAGCGGCTACCTGTCGGCCCTGCGCTACTATGACACCTATCTGTATGTGGGGCGCAGC
76677     ---+---------+---------+---------+---------+---------+---------+------ 76736
          AGGTCGCCGATGGACAGCCGGGACGCGATGATACTGTGGATAGACATACACCCCGCGTCG

c         S   S   G   Y   L   S   A   L   R   Y   Y   D   T   Y   L   Y   V   G   R   S

          GGGAAGCAGGAGAGTGTGCAGCACTTTTACATGCGGTTAGCCGGCTTCTGTGCCTCAACC
76737     ---+---------+---------+---------+---------+---------+---------+------ 76796
          CCCTTCGTCCTCTCACACGTCGTGAAAATGTACGCCAATCGGCCGAAGACACGGAGTTGG

c         G   K   Q   E   S   V   Q   H   F   Y   M   R   L   A   G   F   C   A   S   T

          ACCTGCCTCTACGCGGGTCTCAGGGCAGCCCTGCAGCGGGCCAGGCCGGAGATTGAGAGT
76797     ---+---------+---------+---------+---------+---------+---------+------ 76856
          TGGACGGAGATGCGCCCAGAGTCCCGTCGGGACGTCGCCCGGTCCGGCCTCTAACTCTCA

c         T   C   L   Y   A   G   L   R   A   A   L   Q   R   A   R   P   E   I   E   S

          GACATGGAGGTGTTTGATTACTACTTTGAGCACCTAACCTCCCAGACGGTGTGCTGCTCC
76857     ---+---------+---------+---------+---------+---------+---------+------ 76916
          CTGTACCTCCACAAACTAATGATGAAACTCGTGGATTGGAGGGTCTGCCACACGACGAGG

c         D   M   E   V   F   D   Y   Y   F   E   H   L   T   S   Q   T   V   C   C   S

          ACGCCCTTTATGCGCTTTGCCGGGGTGGAAAACTCCACTCTGGCCAGCTGCATCCTCACC
76917     ---+---------+---------+---------+---------+---------+---------+------ 76976
          TGCGGGAAATACGCGAAACGGCCCCACCTTTTGAGGTGAGACCGGTCGACGTAGGAGTGG
```


- 5 -
T P E M R E A G V E N S T L A S C I L T
ACCCCCGACCTCAGCTCCGAGTGGGACGTGACCCAGGCCCTCTATAGGCACCTGGGGCGC
76977
77036
TGGGGGCTGGAGTCGAGGCTCACCCTGCACTGGGTCCGGGAGATATCCGTGGACCCCGCG
T P D L S S E W D V T Q A L Y R H L G R
TACCTCTTTCAGCGAGCCGGGGTGGGTGTAGGGGTGACGGGGGCTGGCCAGGATGGGAAA
77037
77096
ATGGAGAAAGTCGCTCGGCCCCACCCACATCCCCACTGCCCCCGACCGGTCCTACCCTTT
Y L E Q R A G V G V G V T G A G Q D G K
CACATCAGCCTCCTGATGAGGATGATCAACAGCCACGTGGAGTACCACAACTATGGCTGC
77097
77156
GTGTAGTCGGAGGACTACTCCTACTAGTTGTCGGTGCACCTCATGGTGTTGATACCGACG
H I S L L M R M I N S H V E Y H N Y G C
AAGAGGCCGGTCAGTGTGGCGGCCTACATGGAGCCCTGGCACAGCCAGATTTTCAAGTTT
77157
77216
TTCTCCGGCCAGTCACACCGCCGGATGTACCTCGGGACCGTGTCGGTCTAAAAGTTCAAA
K R P V S V A A Y M E P W H S Q I E K E
TTGGAAACGAAGCTGCCGGAGAACCACGAGAGGTGCCCGGGCATCTTTACGGGGCTCTTT
77217
77276
AACCTTTGCTTCGACGGCCTCTTGGTGCTCTCCACGGGCCCGTAGAAATGCCCCGAGAAA
L E T K L P E N H E R C P G I E T G L E
GTCCCCGAGCTCTTCTTCAAGCTTTTTAGGGACACGCCCTGGTCGGACTGGTACCTGTTT
77277
77336
CAGGGGCTCGAGAAGAAGTTCGAAAAATCCCTGTGCGGGACCAGCCTGACCATGGACAAA
V P E L E E K L E R D T P W S D W Y L E
GACCCCAAGGACGCCGGGGACCTGGAGAGGCTCTACGGGGAGGAGTTTGAGCGCGAGTAC
77337
77396
CTGGGGTTCCTGCGGCCCCTGGACCTCTCCGAGATGCCCCTCCTCAAACTCGCGCTCATG
D P K D A G D L E R L Y G E E F E R E Y
TATCGGCTGGTGACAGCGGGCAAGTTTTGTGGGCGGGTCTCCATCAAGTCCCTGATGTTC
77397
77456
ATAGCCGACCACTGTCGCCCGTTCAAAACACCCGCCCAGAGGTAGTTCAGGGACTACAAG
Y R L V T A G K F C G R V S I K S L M F
TCTATCGTCAACTGCGCCGTCAAGGCCGGCAGCCCCTTCATCCTTTTGAAGGAGGCCTGC
77457
77516
AGATAGCAGTTGACGCGGCAGTTCCGGCCGTCGGGGAAGTAGGAAAACTTCCTCCGGACG
S I V N C A V K A G S P F I L L K E A C
AACGCCCACTTTTGGCGCGACCTGCAGGGCGAGGCCATGAACGCCGCCAACCTGTGCGCC
77517
77576
TTGCGGGTGAAAACCGCGCTGGACGTCCCGCTCCGGTACTTGCGGCGGTTGGACACGCGG
N A H E W R D L Q G E A M N A A N L C A
GAGGTGCTGCAGCCCTCGAGGAAGTCTGTGGCCACCTGCAATCTGGCCAACATCTGCCTC

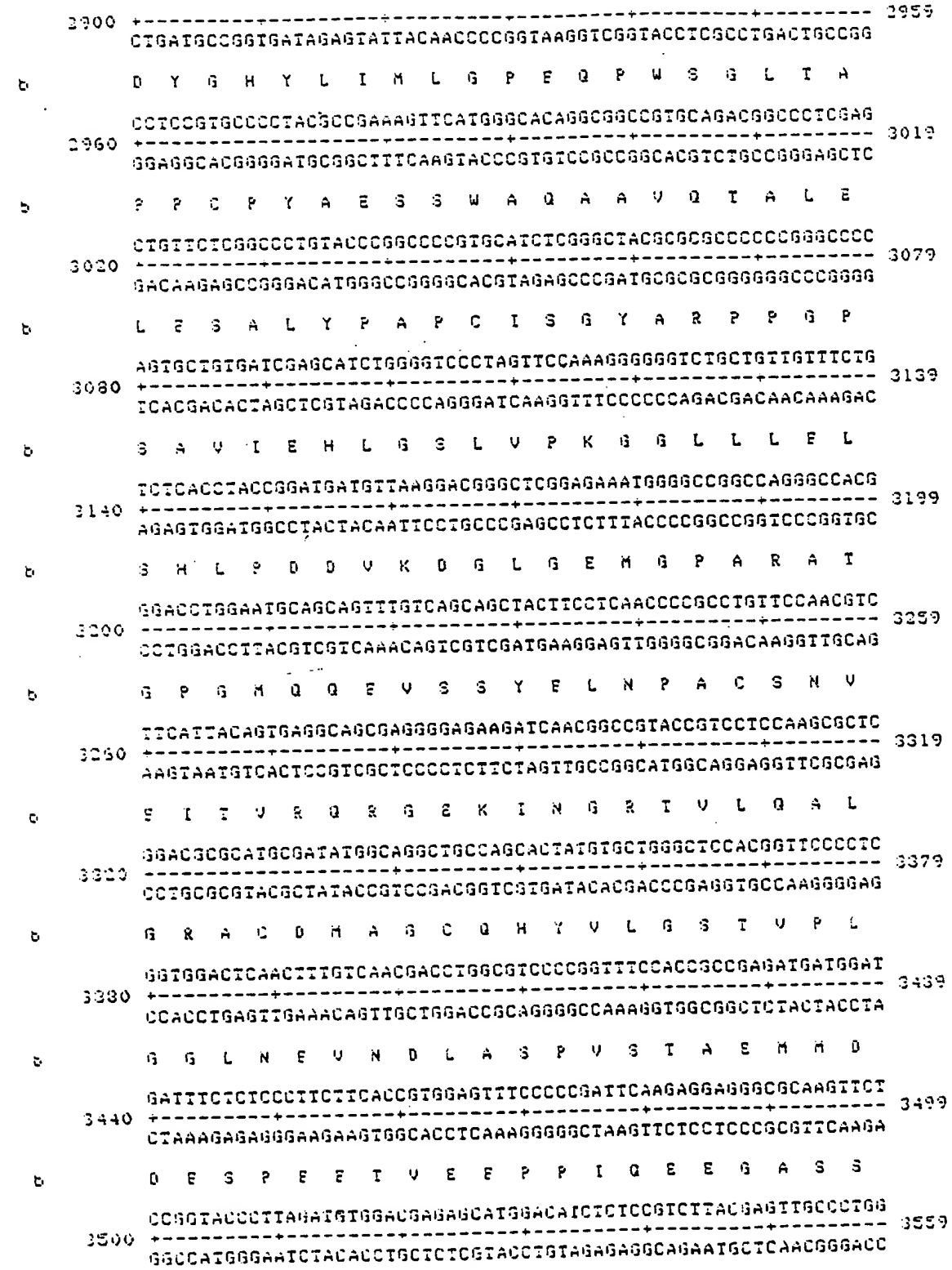

77577
77636
CTCCACGACGTCGGGAGCTCCTTCAGACACCGGTGGACGTTAGACCGGTTGTAGACGGAG
E V L Q P S R K S V A I C N L A N I C L
CCGCGCTGCCTGGTGAATGCGCCTCTGGCGGTGCGGGCACAGCGGGCCGACACGCAGGGG
77637
77696
GGCGCGACGGACCACTTACGCGGAGACCGCCACGCCCGTGTCGCCCGGCTGTGCGTCCCC
P R C L V N A P L A V R A Q R A D I Q G
GATGAACTCCTGCTGGCCCTCCCTCGACTCTCAGTCACCCTACCTGGAGAGGGGGCAGTC
77697
77756
CTACTTGAGGACGACCGGGAGGGAGCTGAGAGTCAGTGGGATGGACCTCTCCCCCGTCAG
D E L L L A L P R L S V T L P G E G A V
GGTGATGGATTCTCGCTAGCCCGCCTCAGAGATGCCACCCAGTGTGCCACCTTTGTGGTG
77757
77816
CCACTACCTAAGAGCGATCGGGCGGAGTCTCTACGGTGGGTCACACGGTGGAAACACCAC
G D G F S L A R L R D A T Q C A T F V V
GCCTGCTCCATTCTTCAGGGATCCCCCACTTATGATTCCAGGGATATGGCCTCCATGGGC
77817
77876
CGGACGAGGTAAGAAGTCCCTAGGGGGTGAATACTAAGGTCCCTATACCGGAGGTACCCG
A C S I L Q G S P T Y D S R D M A S M G
CTCGGGGTGCAGGGCCTGGCCGATGTCTTTGCGGACCTGGGCTGGCAGTACACTGACCCT
77877
77936
GAGCCCCACGTCCCGGACCGGCTACAGAAACGCCTGGACCCGACCGTCATGTGACTGGGA
L G V Q G L A D V F A D L G W Q Y T D P
CCCTCTCGCTCGTTAAACAAGGAAATATTCGAACATATGTACTTTACGGCCCTCTGCACC
77937
77996
GGGAGAGCGAGCAATTTGTTCCTTTATAAGCTTGTATACATGAAATGCCGGGAGACGTGG
P S R S L N K E I F E H M Y F T A L C T
AGTAGTCTGATTGGACTTCACACCAGGAAGATTTTTCCGGGTTTCAAACAGAGCAAGTAT
77997
78056
TCATCAGACTAACCTGAAGTGTGGTCCTTCTAAAAAGGCCCAAAGTTTGTCTCGTTCATA
S S L I G L H T R K I F P G E K Q S K (
GCCGGGGGGTGGTTTCACTGGCACGATTGGGCAGGAACAGACCTTTCTATTCCCAGGGAA
78057
78116
CGGCCCCCCACCAAAGTGACCGTGCTAACCCGTCCTTGTCTGGAAAGATAAGGGTCCCTT
A G G W F H W H D W A G T D L S I P R E
ATTTGGTCTCGCCTCTCTGAACGCATTGTGAGGGATGGGCTTTTCAATTCACAGTTTATC
78117
78176
TAAACCAGAGCGGAGAGACTTGCGTAACACTCCCTACCCGAAAAGTTAAGTGTCAAATAG
I W S R L S E R I V R D G L F N S Q F I
GCCCTGATGCCCACCTCAGGCTGTGCCCAGGTGACGGGCTGTTCGGACGCCTTCTACCCC
78177
78236
CGGGACTACGGGTGGAGTCCGACACGGGTCCACTGCCCGACAAGCCTGCGGAAGATGGGG

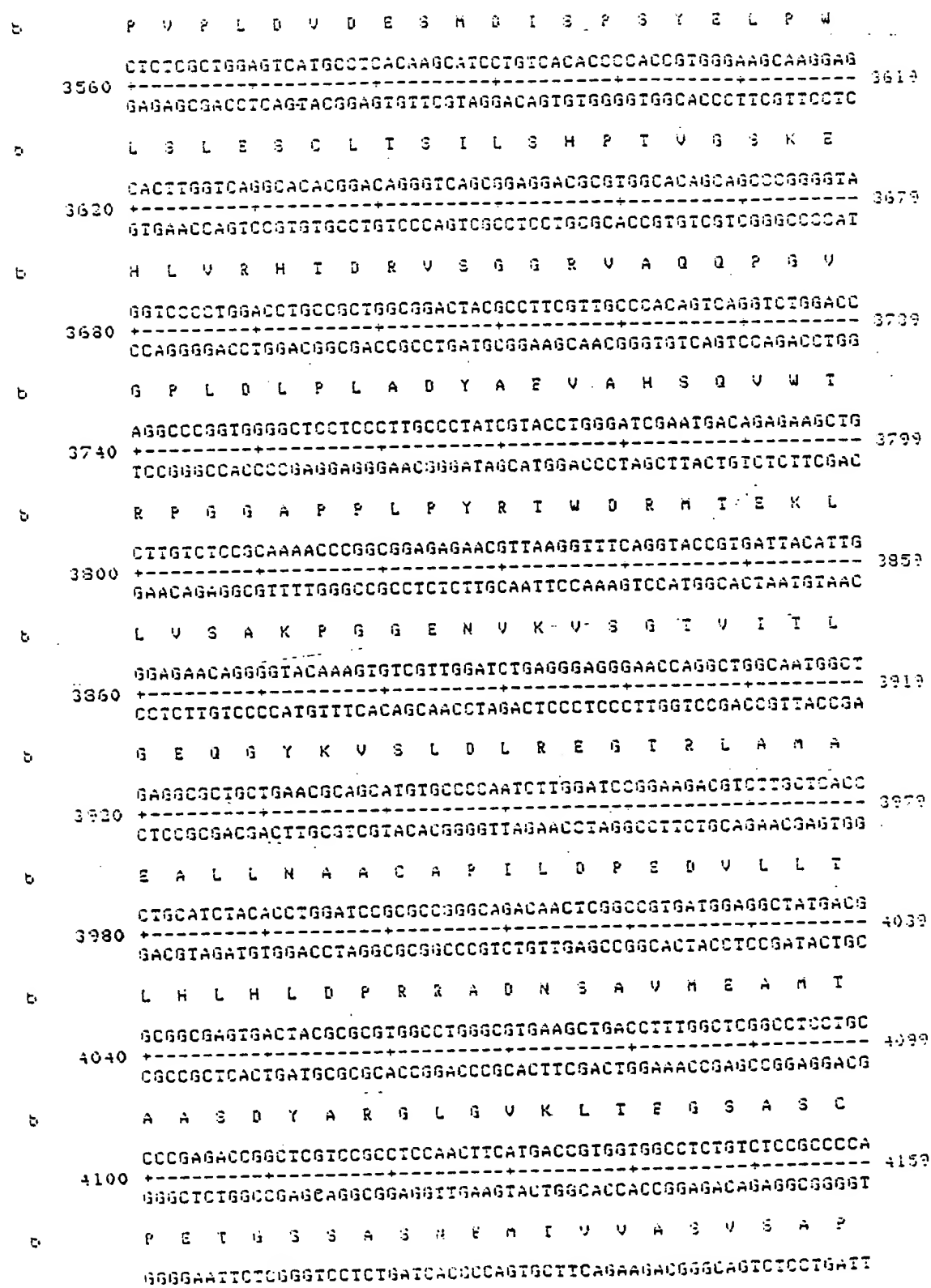
```
c       A  L  M  P  T  S  G  C  A  Q  V  I  G  C  S  D  A  E  Y  P

        ITCTATGCCAATGCGTCCACCAAGGTCACCAACAAGGAGGAGGCCCTTAGGCCAAACCGG
78237   ---+---------+---------+---------+---------+---------+------ 78296
        AAGATACGGTTACGCAGGTGGTTCCAGTGGTTGTTCCTCCTCCGGGAATCCGGTTTGGCC

c       E  Y  A  N  A  S  I  K  V  I  N  K  E  E  A  L  R  P  N  R

        TCTTTTTGGCGTCATGTGCGTCTGGATGACAGGGAAGCTTTGAATCTTGTCGGGGGCCGT
78297   ---+---------+---------+---------+---------+---------+------ 78356
        AGAAAAACCGCAGTACACGCAGACCTACTGTCCCTTCGAAACTTAGAACAGCCCCCGGCA

c       S  E  W  R  H  V  R  L  D  D  R  E  A  L  N  L  V  G  G  R

        GTCTCCTGCCTCCCGGAGGCTCTGCGGCAGCGCTACCTGCGTTTCCAAACGGCCTTTGAT
78357   ---+---------+---------+---------+---------+---------+------ 78416
        CAGAGGACGGAGGGCCTCCGAGACGCCGTCGCGATGGACGCAAAGGTTTGCCGGAAACTA

c       V  S  C  L  P  E  A  L  R  Q  R  Y  L  R  E  Q  T  A  E  D

        TACAACCAGGAGGACCTGATTCAGATGTCCCGGGACAGGGCCCCCTTTGTGGACCAGAGC
78417   ---+---------+---------+---------+---------+---------+------ 78476
        ATGTTGGTCCTCCTGGACTAAGTCTACAGGGCCCTGTCCCGGGGGAAACACCTGGTCTCG

c       Y  N  Q  E  D  L  I  Q  M  S  R  D  R  A  P  E  V  D  Q  S

        CAATCTCACAGCCTGTTTTTGCGTGAGGAAGATGCCGCGCGGGCCAGCACGCTAGCCAAC
78477   ---+---------+---------+---------+---------+---------+------ 78536
        GTTAGAGTGTCGGACAAAAACGCACTCCTTCTACGGCGCGCCCGGTCGTGCGATCGGTTG

c       Q  S  H  S  L  E  L  R  E  E  D  A  A  R  A  S  T  L  A  N

        CTACTGGTGCGCAGCTACGAGCTGGGCCTGAAGACTATCATGTACTATTGTCGCATTGAG
78537   ---+---------+---------+---------+---------+---------+------ 78596
        GATGACCACGCGTCGATGCTCGACCCGGACTTCTGATAGTACATGATAACAGCGTAACTC

c       L  L  V  R  S  Y  E  L  G  L  K  T  I  M  Y  Y  C  R  I  E

        AAGGCCGCCGATCTGGGGGTGATGGAGTGTAAGGCCAGCGCGGCTCTGTCGGTGCCGCGG
78597   ---+---------+---------+---------+---------+---------+------ 78656
        TTCCGGCGGCTAGACCCCCACTACCTCACATTCCGGTCGCGCCGAGACAGCCACGGCGCC

c       K  A  A  D  L  G  V  M  E  C  K  A  S  A  A  L  S  V  P  R

        GAGGAACAGAATGAGCGGAGTCCCGCTGAGCAGATGCCGCCTCGTCCCATGGAACCGGCG
78657   ---+---------+---------+---------+---------+---------+------ 78716
        CTCCTTGTCTTACTCGCCTCAGGGCGACTCGTCTACGGCGGAGCAGGGTACCTTGGCCGC

c       E  E  Q  N  E  R  S  P  A  E  Q  M  P  P  R  P  M  E  P  A

        CAGGTTGCGGGGCCGGTTGACATCATGAGCAAGGGCCCAGGGGAGGGACCAGGTGGGTGG
78717   ---+---------+---------+---------+---------+---------+------ 78776
        GTCCAACGCCCCGGCCAACTGTAGTACTCGTTCCCGGGTCCCCTCCCTGGTCCACCCACC

c       Q  V  A  G  P  V  D  I  M  S  K  G  P  G  E  G  P  G  G  W

        TGTGTGCCCGGGGGATTGGAAGTGTGCTATAAGTACCGTCAGCTCTTCTCAGAGGATGAT
78777   ---+---------+---------+---------+---------+---------+------ 78836
        ACACACGGGCCCCCTAACCTTCACACGATATTCATGGCAGTCGAGAAGAGTCTCCTACTA

c       C  V  P  G  G  L  E  V  C  Y  K  Y  R  Q  L  F  S  E  D  D

        CTGTTGGAGACTGACGGTTTTTACTGAACGAGCCTGTGAATCTTGCCAATAAACGTTTATT
```

```
78837 ---+----------+----------+----------+----------+---------+------ 78896
      GACAACCTCTGACTGCCAAAATGACTTGCTCGGACACTTAGAACGGTTATTTGCAAATAA

      L  L  E  T  D  G  F  T  E  R  A  C  E  S  C  Q  *

      GCCATGTCCAAGTTGTTG
78897 ---+----------+---- 78914
      CGGTACAGGTTCAACAAC
```


ITAGAGACACTATCACGTGTAACTTGACGTGCAAGGATGGAAGAGAGGGGCAGGGAAACG
1700
AATCTCTGTGATAGTGCACATTGAACTGCACGTTCCTACCTTCTCTCCCCGTCCCTTTGC
M E E R G R E T
CAAATGCCGGTTGCCCGGTATGGGGGCCCGTTTATTATGGTAAGGCTCTTCGGGCAAGAT
1760
GTTTACGGCCAACGGGCCATACCCCCGGGCAAATAATACCATTCCGAGAAGCCCGTTCTA
Q M P V A R Y G G P E I M V R L E G Q D
GGAGAGGCAAACATACAGGAGGAAAGGCTATATGAGCTACTCTCTGACCCACGCTCCGCG
1820
CCTCTCCGTTTGTATGTCCTCCTTTCCGATATACTCGATGAGAGACTGGGTGCGAGGCGC
G E A N I Q E E R L Y E L L S D P R S A
CTCGGCCTAGACCCGGGGCCCCTGATTGCTGAGAACCTGCTGCTAGTGGCGCTGCGTGGC
1880
GAGCCGGATCTGGGCCCCGGGGACTAACGACTCTTGGACGACGATCACCGCGACGCACCG
L G L D P G P L I A E N L L L V A L R G
ACCAACAACGATCCCAGGCCTCAGCGTCAGGAGAGGGCCAGAGAACTGGCCCTCGTTGGC
1940
TGGTTGTTGCTAGGGTCCGGAGTCGCAGTCCTCTCCCGGTCTCTTGACCGGGAGCAACCG
I N N D P R P G R Q E R A R E L A L V G
ATTCTACTAGGAAACGGCGAGCAGGGTGAACACTTGGGCACGGAGAGTGCCCTGGAGGCC
2000
TAAGATGATCCTTTGCCGCTCGTCCCACTTGTGAACCCGTGCCTCTCACGGGACCTCCGG
I L L G N G E Q G E H L G T E S A L E A
ICAGGCAACAACTATGTGTATGCCTACGGACCAGACTGGATGGCAAGGCCTTCCACATGG
2060
AGTCCGTTGTTGATACACATACGGATGCCTGGTCTGACCTACCGTTCCGGAAGGTGTACC
S G N N Y V Y A Y G P D W M A R P S T W
TCCGCGGAAATCCAGCAATTCCTGCGACTCCTGGGCGCCACGTACGTGCTTCGCGTGGAG
2120
AGGCGCCTTTAGGTCGTTAAGGACGCTGAGGACCCGCGGTGCATGCACGAAGCGCACCTC
S A E I Q Q E L R L L G A T Y V L R V E
ATGGGCAGGCAGTTTGGCTTCGAGGTGCATAGAAGCCGGCCCTCCTTCCGTCAGTTCCAG
2180
TACCCGTCCGTCAAACCGAAGCTCCACGTATCTTCGGCCGGGAGGAAGGCAGTCAAGGTC
M G R Q E G E E V H R S R P S F R Q F Q
GCCATCAATCACCTTGTCCTGTTTGACAACGCCCTTCGCAAGTACGATTCCGGCCAGGTG
2240
CGGTAGTTAGTGGAACAGGACAAACTGTTGCGGGAAGCGTTCATGCTAAGGCCGGTCCAC

```
b       A  I  N  H  L  V  L  E  D  N  A  L  R  K  Y  D  S  G  Q  V

        GCGGCGGGCTTCCAGAGGGCCCTTCTGGTGGCCGGGCCAGAGACCGCTGACACGAGGCCG
2300    +---------+---------+---------+---------+---------+---------+---------  2359
        CGCCGCCCGAAGGTCTCCCGGGAAGACCACCGGCCCGGTCTCTGGCGACTGTGCTCCGGC

b       A  A  G  E  Q  R  A  L  L  V  A  G  P  E  T  A  D  T  R  P

        GACCTCCGCAAGCTGAATGAGTGGGTGTTTGGTGGCAGGGCTGCTGGTGGCAGACAGCTG
2360    +---------+---------+---------+---------+---------+---------+---------  2419
        CTGGAGGCGTTCGACTTACTCACCCACAAACCACCGTCCCGACGACCACCGTCTGTCGAC

b       D  L  R  K  L  N  E  W  V  E  G  G  R  A  A  G  G  R  Q  L

        GCCGACGAGCTAAAGATCGTGTCCGCGCTGCGAGACACTTACTCGGGCCACTTGGTCCTT
2420    +---------+---------+---------+---------+---------+---------+---------  2479
        CGGCTGCTCGATTTCTAGCACAGGCGCGACGCTCTGTGAATGAGCCCGGTGAACCAGGAA

b       A  D  E  L  K  I  V  S  A  L  R  D  T  Y  S  G  H  L  V  L

        CAGCCCACGGAGACCCTTGACACATGGAAGGTGTTGAGCAGGGACACACGAACCGCTCAT
2480    +---------+---------+---------+---------+---------+---------+---------  2539
        GTCGGGTGCCTCTGGGAACTGTGTACCTTCCACAACTCGTCCCTGTGTGCTTGGCGAGTA

b       Q  P  T  E  T  L  D  T  W  K  V  L  S  R  D  T  R  T  A  H

        AGTTTGGAGCACGGATTCATTCATGCCGCGGGGACCATCCAGGCCAACTGCCCACAGCTG
2540    +---------+---------+---------+---------+---------+---------+---------  2599
        TCAAACCTCGTGCCTAAGTAAGTACGGCGCCCCTGGTAGGTCCGGTTGACGGGTGTCGAC

b       S  L  E  H  G  E  I  H  A  A  G  T  I  Q  A  N  C  P  Q  L

        TTTATGAGACGCCAGCACCCCGGCCTCTTTCCCTTCGTTAATGCAATAGCATCATCGCTG
2600    +---------+---------+---------+---------+---------+---------+---------  2659
        AAATACTCTGCGGTCGTGGGGCCGGAGAAAGGGAAGCAATTACGTTATCGTAGTAGCGAC

b       E  M  R  R  Q  H  P  G  L  E  P  E  V  N  A  I  A  S  S  L

        GGCTGGTACTACCAGACCGCCACCGGCCCCGGAGCAGATGCCAGGGCGGCGGCCCGGCGC
2660    +---------+---------+---------+---------+---------+---------+---------  2719
        CCGACCATGATGGTCTGGCGGTGGCCGGGGCCTCGTCTACGGTCCCGCCGCCGGGCCGCG

b       G  W  Y  Y  Q  T  A  T  G  P  G  A  D  A  R  A  A  A  R  R

        CAACAGGCCTTTCAGACCAGGGCGGCGGCTGAATGCCATGCCAAAAGCGGGGTGCCGGTC
2720    +---------+---------+---------+---------+---------+---------+---------  2779
        GTTGTCCGGAAAGTCTGGTCCCGCCGCCGACTTACGGTACGGTTTTCGCCCCACGGCCAG

b       Q  Q  A  E  Q  T  R  A  A  A  E  C  H  A  K  S  G  V  P  V

        GTGGCCGGCTTCTACAGGACCATCAACGCCACGCTCAAGGGAGGAGAGGGCCTACAGCCC
2780    +---------+---------+---------+---------+---------+---------+---------  2839
        CACCGGCCGAAGATGTCCTGGTAGTTGCGGTGCGAGTTCCCTCCTCTCCCGGATGTCGGG

b       V  A  G  E  Y  R  T  I  N  A  T  L  K  G  G  E  G  L  Q  P

        ACTATGTTTAACGGGGAGCTGGGGGCCATCAAGCACCAGGCACTTGACACTGTGAGGTAT
2840    +---------+---------+---------+---------+---------+---------+---------  2899
        TGATACAAATTGCCCCTCGACCCCCGGTAGTTCGTGGTCCGTGAACTGTGACACTCCATA

b       I  M  E  N  G  E  L  G  A  I  K  H  Q  A  L  D  T  V  R  Y

        GACTACGGCCACTATCTCATAATGTTGGGGCCATTCCAGCCATGGAGCGGACTGACGGCC
```


2900
2959
CTGATGCCGGTGATAGAGTATTACAACCCCGGTAAGGTCGGTACCTCGCCTGACTGCCGG
D Y G H Y L I M L G P E Q P W S G L T A
CCTCCGTGCCCCTACGCCGAAAGTTCATGGGCACAGGCGGCCGTGCAGACGGCCCTCGAG
2960
3019
GGAGGCACGGGGATGCGGCTTTCAAGTACCCGTGTCCGCCGGCACGTCTGCCGGGAGCTC
P P C P Y A E S S W A Q A A V Q T A L E
CTGTTCTCGGCCCTGTACCGGCCCCGTGCATCTCGGGCTACGCGCGCCCCCCGGGCCCC
3020
3079
GACAAGAGCCGGGACATGGGCCGGGGCACGTAGAGCCCGATGCGCGCGGGGGGCCCGGGG
L E S A L Y P A P C I S G Y A R P P G P
AGTGCTGTGATCGAGCATCTGGGGTCCCTAGTTCCAAAGGGGGGTCTGCTGTTGTTTCTG
3080
3139
TCACGACACTAGCTCGTAGACCCCAGGGATCAAGGTTTCCCCCCAGACGACAACAAAGAC
S A V I E H L G S L V P K G G L L L F L
TCTCACCTACCGGATGATGTTAAGGACGGGCTCGGAGAAATGGGGCCGGCCAGGGCCACG
3140
3199
AGAGTGGATGGCCTACTACAATTCCTGCCCGAGCCTCTTTACCCCGGCCGGTCCCGGTGC
S H L P D D V K D G L G E M G P A R A T
GGACCTGGAATGCAGCAGTTTGTCAGCAGCTACTTCCTCAACCCCGCCTGTTCCAACGTC
3200
3259
CCTGGACCTTACGTCGTCAAACAGTCGTCGATGAAGGAGTTGGGGCGGACAAGGTTGCAG
G P G M Q Q F V S S Y F L N P A C S N V
TTCATTACAGTGAGGCAGCGAGGGGAGAAGATCAACGGCCGTACCGTCCTCCAAGCGCTC
3260
3319
AAGTAATGTCACTCCGTCGCTCCCCTCTTCTAGTTGCCGGCATGGCAGGAGGTTCGCGAG
S I T V R Q R G E K I N G R T V L Q A L
GGACGCGCATGCGATATGGCAGGCTGCCAGCACTATGTGCTGGGCTCCACGGTTCCCCTC
3320
3379
CCTGCGCGTACGCTATACCGTCCGACGGTCGTGATACACGACCCGAGGTGCCAAGGGGAG
G R A C D M A G C Q H Y V L G S T V P L
GGTGGACTCAACTTTGTCAACGACCTGGCGTCCCCGGTTTCCACCGCCGAGATGATGGAT
3380
3439
CCACCTGAGTTGAAACAGTTGCTGGACCGCAGGGGCCAAAGGTGGCGGCTCTACTACCTA
G G L N E V N D L A S P V S T A E M M D
GATTTCTCTCCCTTCTTCACCGTGGAGTTTCCCCCGATTCAAGAGGAGGGCGCAAGTTCT
3440
3499
CTAAAGAGAGGGAAGAAGTGGCACCTCAAAGGGGGCTAAGTTCTCCTCCCGCGTTCAAGA
D E S P E E T V E E P P I Q E E G A S S
CCGGTACCCTTAGATGTGGACGAGAGCATGGACATCTCTCCGTCTTACGAGTTGCCCTGG
3500
3559
GGCCATGGGAATCTACACCTGCTCTCGTACCTGTAGAGAGGCAGAATGCTCAACGGGACC

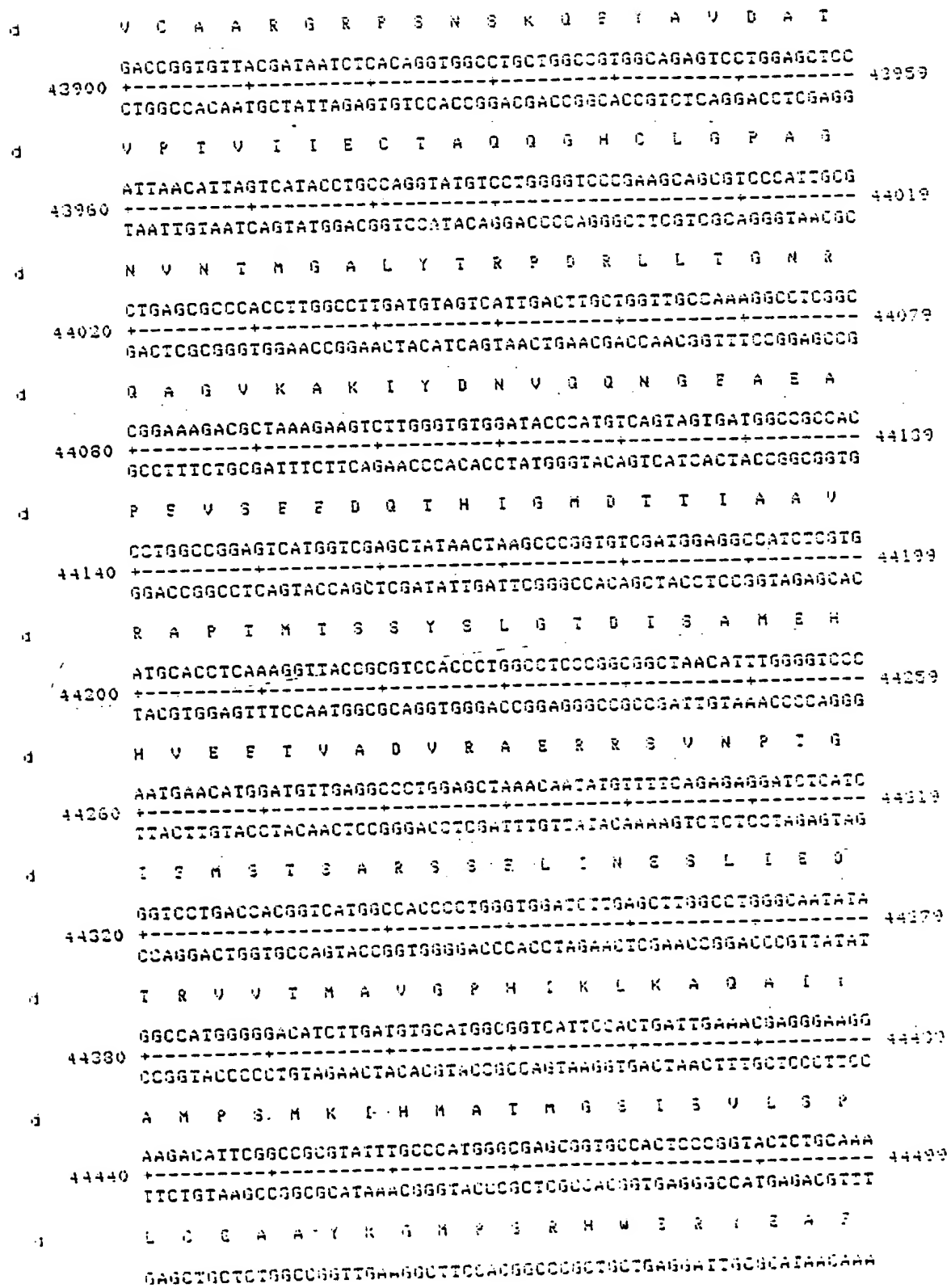

P V P L D V D E S M D I S P S Y E L P W
CTCTCGCTGGAGTCATGCCTCACAAGCATCCTGTCACACCCCACCGTGGGAAGCAAGGAG
3560
3619
GAGAGCGACCTCAGTACGGAGTGTTCGTAGGACAGTGTGGGGTGGCACCCTTCGTTCCTC
L S L E S C L T S I L S H P T V G S K E
CACTTGGTCAGGCACACGGACAGGGTCAGCGGAGGACGCGTGGCACAGCAGCCCGGGGTA
3620
3679
GTGAACCAGTCCGTGTGCCTGTCCCAGTCGCCTCCTGCGCACCGTGTCGTCGGGCCCCAT
H L V R H T D R V S G G R V A Q Q P G V
GGTCCCCTGGACCTGCCGCTGGCGGACTACGCCTTCGTTGCCCACAGTCAGGTCTGGACC
3680
3739
CCAGGGGACCTGGACGGCGACCGCCTGATGCGGAAGCAACGGGTGTCAGTCCAGACCTGG
G P L D L P L A D Y A F V A H S Q V W T
AGGCCCGGTGGGGCTCCTCCCTTGCCCTATCGTACCTGGGATCGAATGACAGAGAAGCTG
3740
3799
TCCGGGCCACCCCGAGGAGGGAACGGGATAGCATGGACCCTAGCTTACTGTCTCTTCGAC
R P G G A P P L P Y R I W D R M T E K L
CTTGTCTCCGCAAAACCCGGCGGAGAGAACGTTAAGGTTTCAGGTACCGTGATTACATTG
3800
3859
GAACAGAGGCGTTTTGGGCCGCCTCTCTTGCAATTCCAAAGTCCATGGCACTAATGTAAC
L V S A K P G G E N V K V S G T V I T L
GGAGAACAGGGGTACAAAGTGTCGTTGGATCTGAGGGAGGGAACCAGGCTGGCAATGGCT
3860
3919
CCTCTTGTCCCCATGTTTCACAGCAACCTAGACTCCCTCCCTTGGTCCGACCGTTACCGA
G E Q G Y K V S L D L R E G I R L A M A
GAGGCGCTGCTGAACGCAGCATGTGCCCCAATCTTGGATCCGGAAGACGTCTTGCTCACC
3920
3979
CTCCGCGACGACTTGCGTCGTACACGGGGTTAGAACCTAGGCCTTCTGCAGAACGAGTGG
E A L L N A A C A P I L D P E D V L L T
CTGCATCTACACCTGGATCCGCGCCGGGCAGACAACTCGGCCGTGATGGAGGCTATGACG
3980
4039
GACGTAGATGTGGACCTAGGCGCGGCCCGTCTGTTGAGCCGGCACTACCTCCGATACTGC
L H L H L D P R R A D N S A V M E A M I
GCGGCGAGTGACTACGCGCGTGGCCTGGGCGTGAAGCTGACCTTTGGCTCGGCCTCCTGC
4040
4099
CGCCGCTCACTGATGCGCGCACCGGACCCGCACTTCGACTGGAAACCGAGCCGGAGGACG
A A S D Y A R G L G V K L T F G S A S C
CCCGAGACCGGCTCGTCCGCCTCCAACTTCATGACCGTGGTGGCCTCTGTCTCCGCCCCA
4100
4159
GGGCTCTGGCCGAGCAGGCGGAGGTTGAAGTACTGGCACCACCGGAGACAGAGGCGGGGT
P E T G S S A S N F M T V V A S V S A P
GGGGAATTCTCGGGTCCTCTGATCACCCAGTGCTTCAGAAGACGGGCAGTCTCCTGATT

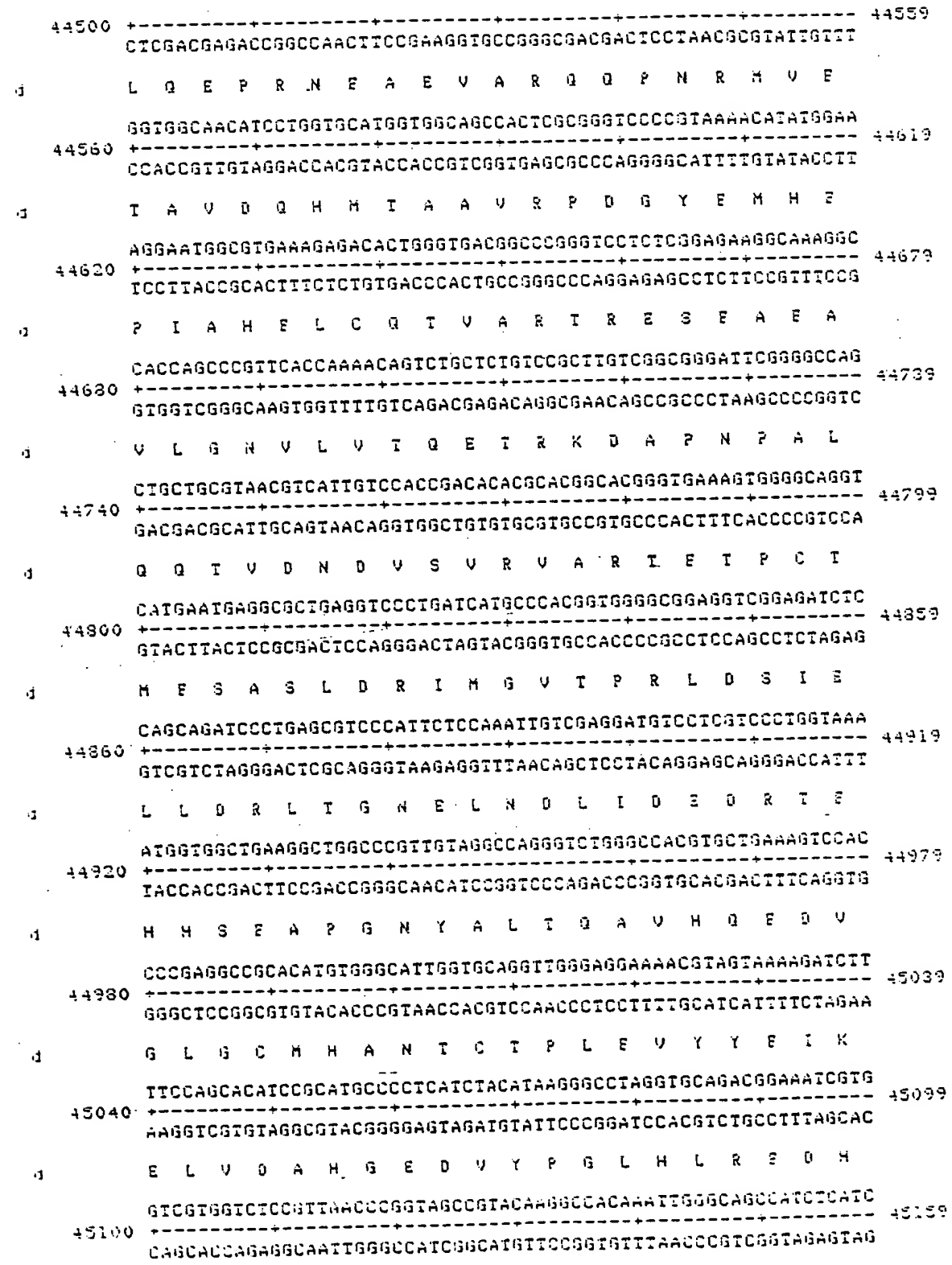
```
4160 +---------+---------+---------+---------+---------+---------+---------+ 4219
     CCCCTTAAGAGCCCAGGAGACTAGTGCGGTCACGAAGTCTTCTGCCCGTCAGAGGACTAA

b        G   E   E   S   G   P   L   I   T   P   V   L   Q   K   T   G   S   L   L   I

     GCGGTGCGTTGCGGGGATGGCAAGATCCAGGGAGGGTCGCTGTTTGAGCAGCTCTTTAGC
4220 +---------+---------+---------+---------+---------+---------+---------+ 4279
     CGCCACGCAACGCCCCTACCGTTCTAGGTCCCTCCCAGCGACAAACTCGTCGAGAAATCG

b        A   V   R   C   G   D   G   K   I   Q   G   G   S   L   F   E   Q   L   F   S

     GACGTGGCCACGACCCCACGGGCACCCGAGGCGTTGTCTCTGAAGAATCTCTTCCGGGCA
4280 +---------+---------+---------+---------+---------+---------+---------+ 4339
     CTGCACCGGTGCTGGGGTGCCCGTGGGCTCCGCAACAGAGACTTCTTAGAGAAGGCCCGT

b        D   V   A   I   I   P   R   A   P   E   A   L   S   L   K   N   L   F   R   A

     GTCCAGCAGCTGGTCAAGAGCGGCATCGTGCTGTCAGGGCATGACATCAGCGACGGGGGC
4340 +---------+---------+---------+---------+---------+---------+---------+ 4399
     CAGGTCGTCGACCAGTTCTCGCCGTAGCACGACAGTCCCGTACTGTAGTCGCTGCCCCCG

b        V   Q   Q   L   V   K   S   G   I   V   L   S   G   H   D   I   S   D   G   G

     CTGGTGACCTGCCTGGTGGAGATGGCCCTGGCCGGGCAGCGGGGAGTGACCATCACTATG
4400 +---------+---------+---------+---------+---------+---------+---------+ 4459
     GACCACTGGACGGACCACCTCTACCGGGACCGGCCCGTCGCCCCTCACTGGTAGTGATAC

b        L   V   T   C   L   V   E   M   A   L   A   G   Q   R   G   V   T   I   T   M

     CCGGTGGCCTCCGACTACCTCCCGGAGATGTTTGCAGAGCACCCCGGCCTGGTGTTTGAG
4460 +---------+---------+---------+---------+---------+---------+---------+ 4519
     GGCCACCGGAGGCTGATGGAGGGCCTCTACAAACGTCTCGTGGGGCCGGACCACAAACTC

b        P   V   A   S   D   Y   L   P   E   M   F   A   E   H   P   G   L   V   F   E

     GTGGAGGAGCGCAGCGTGGGTGAGGTGCTGCAGACCCTGCGCTCCATGAACATGTACCCG
4520 +---------+---------+---------+---------+---------+---------+---------+ 4579
     CACCTCCTCGCGTCGCACCCACTCCACGACGTCTGGGACGCGAGGTACTTGTACATGGGC

b        V   E   E   R   S   V   G   E   V   L   Q   T   L   R   S   M   N   M   Y   P

     GCAGTCCTCGGTCGAGTGGGCGAGCAAGGTCCAGATCAAATGTTTGAGGTGCAGCACGGC
4580 +---------+---------+---------+---------+---------+---------+---------+ 4639
     CGTCAGGAGCCAGCTCACCCGCTCGTTCCAGGTCTAGTTTACAAACTCCACGTCGTGCCG

b        A   V   L   G   R   V   G   E   Q   G   P   D   Q   M   F   E   V   Q   H   G

     CCAGAGACGGTGTTGCGCCAGTCGCTGCGCCTGCTGCTGGGAACCTGGTCATCCTTTGCC
4640 +---------+---------+---------+---------+---------+---------+---------+ 4699
     GGTCTCTGCCACAACGCGGTCAGCGACGCGGACGACGACCCTTGGACCAGTAGGAAACGG

b        P   E   T   V   L   R   Q   S   L   R   L   L   L   G   T   W   S   S   E   A

     AGCGAGCAGTACGAGTGCCTGCGACCAGATCGGATTAACCGGTCCATGCACGTGTCCGAC
4700 +---------+---------+---------+---------+---------+---------+---------+ 4759
     TCGCTCGTCATGCTCACGGACGCTGGTCTAGCCTAATTGGCCAGGTACGTGCACAGGCTG

b        S   E   Q   Y   E   C   L   R   P   D   R   I   N   R   S   M   H   V   S   D

     TACGGCTATAACGAAGCACTGGCAGTCTCCCCGTTGACAGGAAAGAATCTCAGCCCACGC
4760 +---------+---------+---------+---------+---------+---------+---------+ 4819
     ATGCCGATATTGCTTCGTGACCGTCAGAGGGGCAACTGTCCTTTCTTAGAGTCGGGTGCG
```

```
b       Y  G  Y  N  E  A  L  A  V  S  P  L  T  G  K  N  L  S  F  R

        CGGTTGGTGACAGAGCCTGACCCACGATGTCAGGTGGCCGTGCTATGCGCCCCGGGCACC
4820    +---------+---------+---------+---------+---------+---------+--------- 4879
        GCCAACCACTGTCTCGGACTGGGTGCTACAGTCCACCGGCACGATACGCGGGGCCCGTGG

b       R  L  V  T  E  P  D  P  R  C  Q  V  A  V  L  C  A  P  G  T

        AGGGGCCATGAAAGCCTCCTGGCGGCCTTCACGAATGCCGGATGCCTGTGCCGACGGGTG
4880    +---------+---------+---------+---------+---------+---------+--------- 4939
        TCCCCGGTACTTTCGGAGGACCGCCGGAAGTGCTTACGGCCTACGGACACGGCTGCCCAC

b       R  G  H  E  S  L  L  A  A  E  T  N  A  G  C  L  C  R  R  V

        TTCTTTCGCGAGGTTAGGGACAACACGTTCCTCGACAAGTACGTGGGTCTGGCCATCGGA
4940    +---------+---------+---------+---------+---------+---------+--------- 4999
        AAGAAAGCGCTCCAATCCCTGTTGTGCAAGGAGCTGTTCATGCACCCAGACCGGTAGCCT

b       E  E  R  E  V  R  D  N  T  E  L  D  K  Y  V  G  L  A  I  G

        GGAGTTCATGGGGCCAGGGACTCTGCCCTGGCAGGCCGTGCCACCGTGGCGCTGATTAAT
5000    +---------+---------+---------+---------+---------+---------+--------- 5059
        CCTCAAGTACCCCGGTCCCTGAGACGGGACCGTCCGGCACGGTGGCACCGCGACTAATTA

b       G  V  H  G  A  R  D  S  A  L  A  G  R  A  I  V  A  L  I  N

        CGTTTCCCCGCCCTGCGTGACGCTATTCTAAAGTTCCTCAACAGGCCAGATACGTTCTCG
5060    +---------+---------+---------+---------+---------+---------+--------- 5119
        GCAAAGGGGCGGGACGCACTGCGATAAGATTTCAAGGAGTTGTCCGGTCTATGCAAGAGC

b       R  E  P  A  L  R  D  A  I  L  K  E  L  N  R  P  D  T  E  S

        GTGGCCTTGGGGGAGCTGGGGGTGCAAGTTTTGGCTGGCCTGGGGGCCGTGGGGTCAACA
5120    +---------+---------+---------+---------+---------+---------+--------- 5179
        CACCGGAACCCCCTCGACCCCCACGTTCAAAACCGACCGGACCCCCGGCACCCCAGTTGT

b       V  A  L  G  E  L  G  V  Q  V  L  A  G  L  G  A  V  G  S  I

        GATAATCCACCCGCCCCTGGCGTGGAAGTTAATGTCCAGAGATCACCTCTGATTCTGGCC
5180    +---------+---------+---------+---------+---------+---------+--------- 5239
        CTATTAGGTGGGCGGGGACCGCACCTTCAATTACAGGTCTCTAGTGGAGACTAAGACCGG

b       D  N  P  P  A  P  G  V  E  V  N  V  Q  R  S  P  L  I  L  A

        CCCAACGCCTCTGGCATGTTTGAGTCCCGCTGGCTGAACATTAGCATCCCGGCGACCACC
5240    +---------+---------+---------+---------+---------+---------+--------- 5299
        GGGTTGCGGAGACCGTACAAACTCAGGGCGACCGACTTGTAATCGTAGGGCCGCTGGTGG

b       P  N  A  S  G  M  E  E  S  R  W  L  N  I  S  I  P  A  I  I

        AGCTCTGTCATGCTGCGTGGCCTCCGGGGCTGCGTCCTGCCTTGTTGGGTGCAAGGCTCG
5300    +---------+---------+---------+---------+---------+---------+--------- 5359
        TCGAGACAGTACGACGCACCGGAGGCCCCGACGCAGGACGGAACAACCCACGTTCCGAGC

b       S  S  V  M  L  R  G  L  R  G  C  V  L  P  C  W  V  Q  G  S

        TGCCTGGGCCTGCAATTTACTAACCTCGGGATGCCATATGTTTTGCAGAATGCCCACCAG
5360    +---------+---------+---------+---------+---------+---------+--------- 5419
        ACGGACCCGGACGTTAAATGATTGGAGCCCTACGGTATACAAAACGTCTTACGGGTGGTC

b       C  L  G  L  Q  F  T  N  L  G  M  P  Y  V  L  Q  N  A  H  Q

        ATCGCCTGCCACTTCCACAGCAATGGCACGGATGCCTGGCGCTTTGCTATGAATTATCCA
```

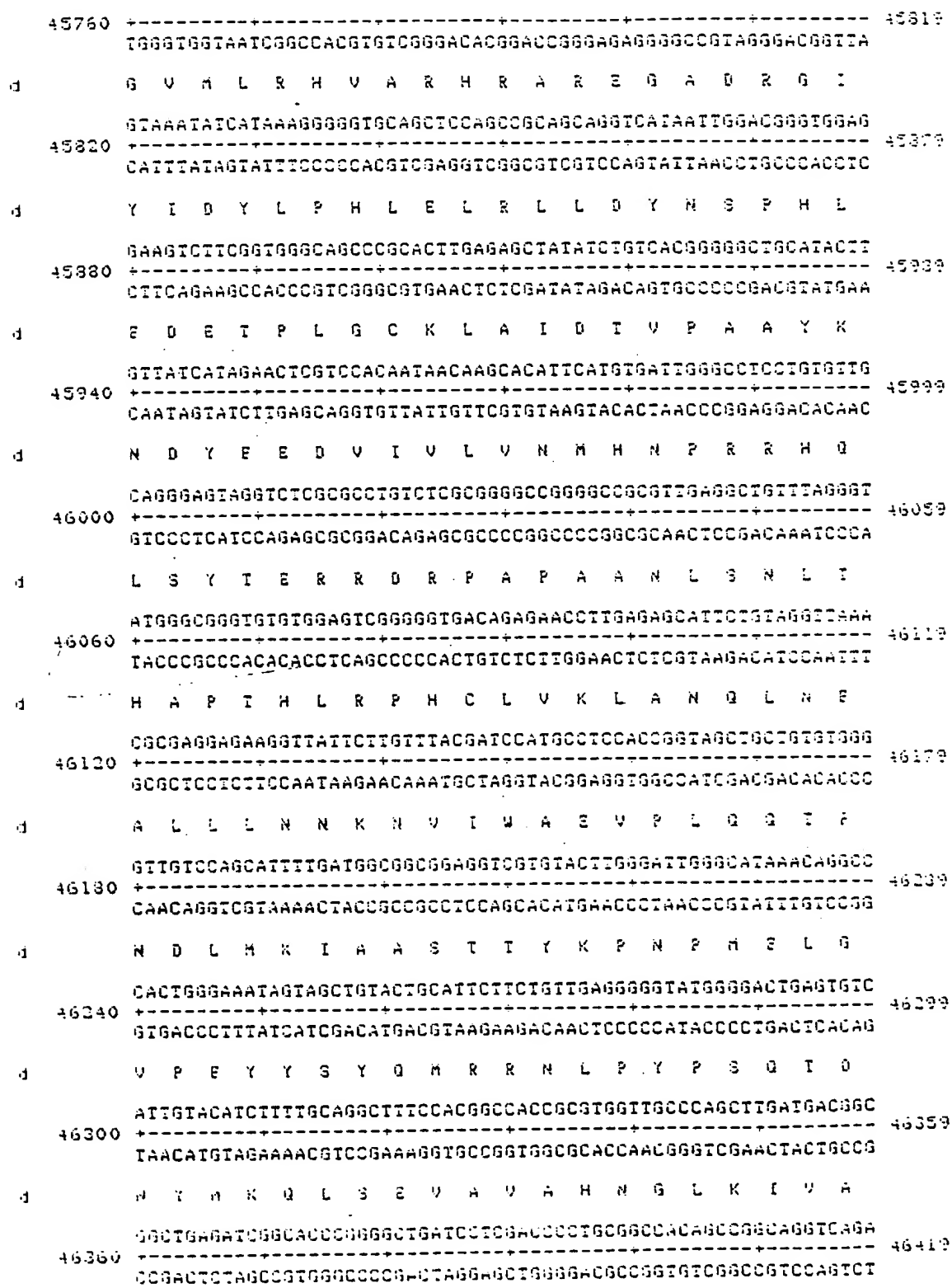
```
5420 +---------+---------+---------+---------+---------+---------+ 5479
     TAGCGGACGGTGAAGGTGTCGTTACCGTGCCTACGGACCGCGAAACGATACTTAATAGGT

b        I   A   C   H   E   H   S   N   G   T   D   A   W   R   E   A   M   N   Y   P

     AGAAACCCCACGGAGCAGGGCAACATTGCAGGGCTCTGTTCACGCGATGGTCGTCATCTG
5480 +---------+---------+---------+---------+---------+---------+ 5539
     TCTTTGGGGTGCCTCGTCCCGTTGTAACGTCCCGAGACAAGTGCGCTACCAGCAGTAGAC

b        R   N   P   T   E   Q   G   N   I   A   G   L   C   S   R   D   G   R   H   L

     GCTCTCCTGTGTGACCCCTCACTTTGTACAGACTTTTGGCAATGGGAGCACATTCCCCCC
5540 +---------+---------+---------+---------+---------+---------+ 5599
     CGAGAGGACACACTGGGGAGTGAAACATGTCTGAAAACCGTTACCCTCGTGTAAGGGGGG

b        A   L   L   C   D   P   S   L   C   T   D   E   W   Q   W   E   H   I   P   P

     GCCTTTGGGCACCCCACGGGGTGCTCCCCCTGGACACTTATGTTTCAAGCAGCTCACCTA
5600 +---------+---------+---------+---------+---------+---------+ 5659
     CGGAAACCCGTGGGGTGCCCCACGAGGGGGACCTGTGAATACAAAGTTCGTCGAGTGGAT

b        A   E   G   H   P   T   G   C   S   P   W   T   L   M   F   Q   A   A   H   L

     TGGTCACTCAGGCACGGTCGCCCCTCCGAGTGACCAGTCAC
5660 +---------+---------+---------+---------+ 5700
     ACCAGTGAGTCCGTGCCAGCGGGGAGGCTCACTGGTCAGTG

b        W   S   L   R   H   G   R   P   S   E   *
```

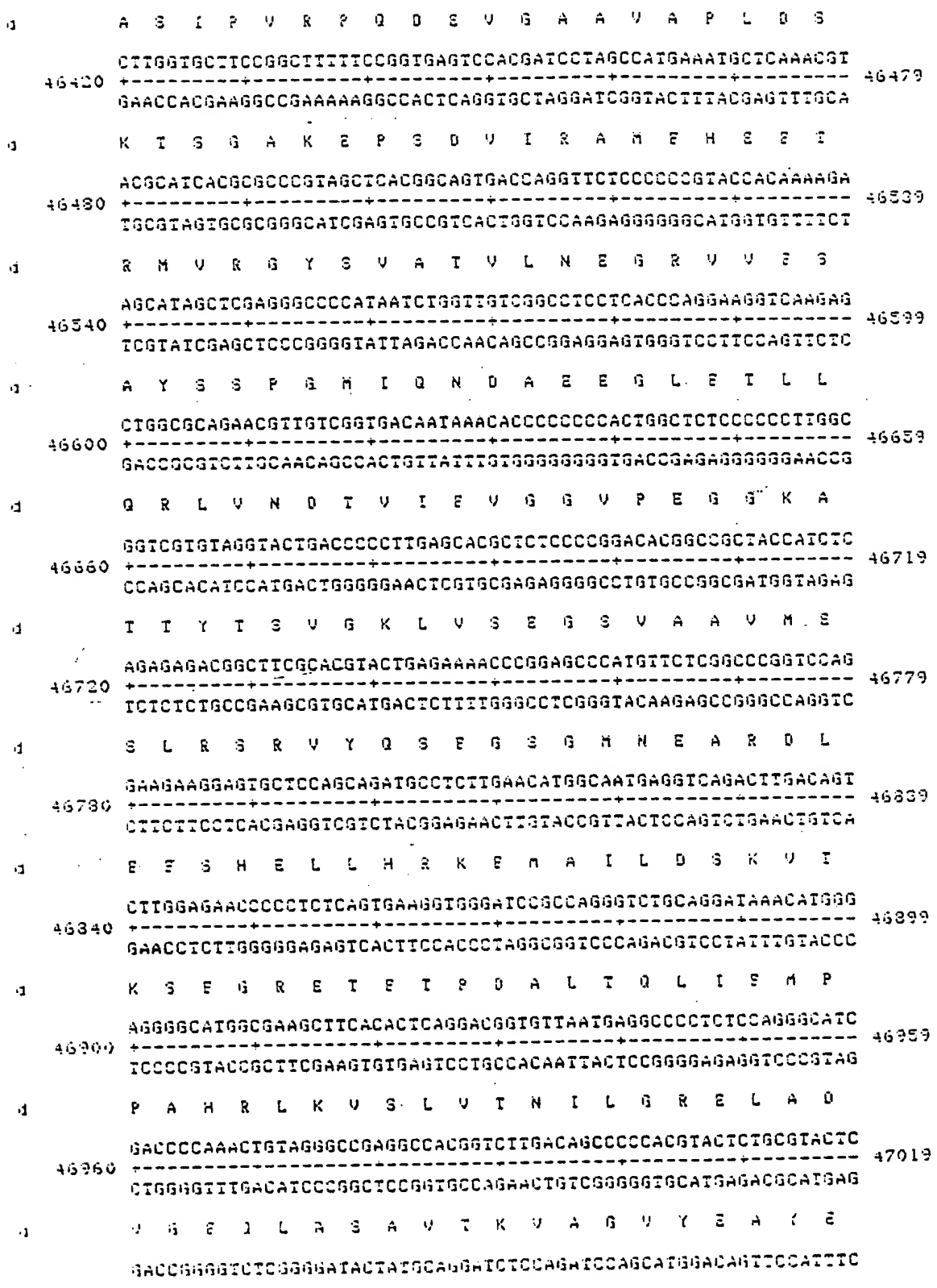
```
          ATCCATCTTTATTGACAATTATCAAAAAACCACCTTATTTCCAAACTTTAATATTCTTCG
   43300  ---------+---------+---------+---------+---------+---------  43359
          TAGGTAGAAATAACTGTTAATAGTTTTTTGGTGGAATAAAGGTTTGAAATTATAAGAAGC

                               A  A  E  V  V  K  N  G  E  K  L  I  R  R

          TACCGGCGCCACCTCTTCAATTATATAGTGTCCGTAATGGATGGGGGCGTGGGTCTGTTT
   43360  +---------+---------+---------+---------+---------+---------  43419
          ATGGCCGCGGTGGAGAAGTTAATATATCACAGGCATTACCTACCCCCGCACCCAGACAAA

          V  P  A  V  E  E  I  I  Y  H  G  Y  H  I  P  A  H  T  Q  K

          GACAGACATAAACTCATCGATGAGTGCCCGGGAGGAGGCTGAGAGTGCGGGGAATGCCTC
   43420  +---------+---------+---------+---------+---------+---------  43479
          CTGTCTGTATTTGAGTAGCTACTCACGGGCCCTCCTCCGACTCTCACGCCCCTTACGGAG

          V  S  M  E  E  D  I  L  A  R  S  S  A  S  L  A  P  E  A  E

          CTGCAGAAAGCTGCAGGGCTGCTCCAGAAACACGTCAGTGCCAGCAATCACTACAAACTG
   43480  +---------+---------+---------+---------+---------+---------  43539
          GACGTCTTTCGACGTCCCGACGAGGTCTTTGTGCAGTCACGGTCGTTAGTGATGTTTGAC

          Q  L  E  S  C  P  Q  E  L  E  V  D  I  G  A  I  V  V  E  Q

          CACCTCTGTGTTGCTGGTGGCTGGGTGCCCTCCAAGTCGCTGGCTGTACTCGTTGACCAT
   43540  +---------+---------+---------+---------+---------+---------  43599
          GTGGAGACACAACGACCACCGACCCACGGGAGGTTCAGCGACCGACATGAGCAACTGGTA

          V  E  T  N  S  T  A  P  H  G  G  L  R  Q  S  Y  E  N  V  M

          GTTGTAGAGTCCCCTGTTGTTGCGCAGAAGCTCCTCCTTGTTGAAAAATGCCCGGCAGGG
   43600  +---------+---------+---------+---------+---------+---------  43659
          CAACATCTCAGGGGACAACAACGCGTCTTCGAGGAGGAACAACTTTTTACGGGCCGTCCC

          N  Y  L  G  R  N  N  R  L  L  E  E  K  N  E  E  A  R  C  P

          GCTGTAGAGGCCCGGGACGGCCGTCTGGCGATAGGAGGAGTTGTACATGATGTCACCCAG
   43660  +---------+---------+---------+---------+---------+---------  43719
          CGACATCTCCGGGCCCTGCCGGCAGACCGCTATCCTCCTCAACATGTACTACAGTGGGTC

          S  Y  L  G  P  V  A  T  Q  R  Y  S  S  N  Y  M  I  D  G  L

          AGAACCCAGCTGAGATGCCCAGGGATTCACAGTGCTCCGGTATTCATAGGCGGCATCCGG
   43720  +---------+---------+---------+---------+---------+---------  43779
          TCTTGGGTCGACTCTACGGGTCCCTAAGTGTCACGAGGCCATAAGTATCCGCCGTAGGCC

          S  G  L  Q  S  A  W  P  N  V  T  S  R  Y  E  Y  A  A  D  P

          GCGAGAATGGTCATAGATGAGCCCCTCGGCAACCTCCTGATTGTAGTTTTCACAGGAGAC
   43780  +---------+---------+---------+---------+---------+---------  43839
          CGCTCTTACCAGTATCTACTCGGGGAGCCGTTGGAGGACTAACATCAAAAGTGTCCTCTG

          R  S  H  D  Y  I  L  G  S  A  V  E  Q  N  Y  N  E  C  S  V

          CACACAGGCGGCCCGTCCCCTTGGAGAGTTGGACTTTTGAAAATAAGCCACGTCTGCCGT
   43840  +---------+---------+---------+---------+---------+---------  43899
          GTGTGTCCGCCGGGCAGGGGAACCTCTCAACCTGAAAACTTTTATTCGGTGCAGACGGCA
```


V C A A R G R P S N S K Q F I A V D A I
GACCGGIGTTACGATAAICTCACAGGTGGCCTGCTGGCCGTGGCAGAGTCCTGGAGCTCC
43900
CTGGCCACAATGCTATTAGAGTGTCCACCGGACGACCGGCACCGTCTCAGGACCTCGAGG
43959
V P T V I I E C T A Q Q G H C L G P A G
ATTAACATTAGTCATACCTGCCAGGTATGTCCTGGGGTCCCGAAGCAGCGTCCCATTGCG
43960
TAATTGTAATCAGTATGGACGGTCCATACAGGACCCCAGGGCTTCGTCGCAGGGTAACGC
44019
N V N T M G A L Y T R P D R L L T G N R
CTGAGCGCCCACCTTGGCCTTGATGTAGTCATTGACTTGCTGGTTGCCAAAGGCCTCGGC
44020
GACTCGCGGGTGGAACCGGAACTACATCAGTAACTGAACGACCAACGGTTTCCGGAGCCG
44079
Q A G V K A K I Y D N V Q Q N G E A E A
CGGAAAGACGCTAAAGAAGTCTTGGGTGTGGATACCCATGTCAGTAGTGATGGCCGCCAC
44080
GCCTTTCTGCGATTTCTTCAGAACCCACACCTATGGGTACAGTCATCACTACCGGCGGTG
44139
P S V S E E D Q I H I G M D I I I A A V
CCTGGCCGGAGTCATGGTCGAGCTATAACTAAGCCCGGTGTCGATGGAGGCCATCTCGTG
44140
GGACCGGCCTCAGTACCAGCTCGATATTGATTCGGGCCACAGCTACCTCCGGTAGAGCAC
44199
R A P I M I S S Y S L G I D I S A M E H
ATGCACCTCAAAGGTTACCGCGTCCACCCTGGCCTCCCGGCGGCTAACATTTGGGGTCCC
44200
TACGTGGAGTTTCCAATGGCGCAGGTGGGACCGGAGGGCCGCCGATTGTAAACCCCAGGG
44259
H V E E I V A D V R A E R R S V N P I G
AATGAACATGGATGTTGAGGCCCTGGAGCTAAACAATATGTTTTCAGAGAGGATCTCATC
44260
TTACTTGTACCTACAACTCCGGGACCTCGATTTGTTATACAAAAGTCTCTCCTAGAGTAG
44319
I S M S I S A R S S E L I N E S L I E D
GGTCCTGACCACGGTCATGGCCACCCCTGGGTGGATCTTGAGCTTGGCCTGGGCAATATA
44320
CCAGGACTGGTGCCAGTACCGGTGGGGACCCACCTAGAACTCGAACCGGACCCGTTATAT
44379
I R V V I M A V G P H I K L K A Q A I I
GGCCATGGGGGACATCTTGATGTGCATGGCGGTCATTCCACTGATTGAAACGAGGGAAGG
44380
CCGGTACCCCCTGTAGAACTACACGTACCGCCAGTAAGGTGACTAACTTTGCTCCCTTCC
44439
A M P S M K D H M A I M G S I S V L S P
AAGACATTCGGCCGCGTATTTGCCCATGGGCGAGCGGTGCCACTCCCGGTACTCTGCAAA
44440
TTCTGTAAGCCGGCGCATAAACGGGTACCCGCTCGCCACGGTGAGGGCCATGAGACGTTT
44499
L C G A A Y K G M P S R H W I R I E A P
GAGCTGCTCTGGCCGGTTGAAGGCTTCCACGGCCCGCTGCTGAGGATTGCGCATAAGAAA


44500
44559
CTCGACGAGACCGGCCAACTTCCGAAGGTGCCGGGCGACGACTCCTAACGCGTATTGTTT
L Q E P R N E A E V A R Q Q P N R M V E
GGTGGCAACATCCTGGTGCATGGTGGCAGCCACTCGCGGGTCCCCGTAAAACATATGGAA
44560
44619
CCACCGTTGTAGGACCACGTACCACCGTCGGTGAGCGCCCAGGGGCATTTTGTATACCTT
T A V D Q H M I A A V R P D G Y E M H E
AGGAATGGCGTGAAAGAGACACTGGGTGACGGCCCGGGTCCTCTCGGAGAAGGCAAAGGC
44620
44679
TCCTTACCGCACTTTCTCTGTGACCCACTGCCGGGCCCAGGAGAGCCTCTTCCGTTTCCG
P I A H E L C Q T V A R T R E S E A E A
CACCAGCCCGTTCACCAAAACAGTCTGCTCTGTCCGCTTGTCGGCGGGATTCGGGGCCAG
44680
44739
GTGGTCGGGCAAGTGGTTTTGTCAGACGAGACAGGCGAACAGCCGCCCTAAGCCCCGGTC
V L G N V L V T Q E T R K D A P N P A L
CTGCTGCGTAACGTCATTGTCCACCGACACACGCACGGCACGGGTGAAAGTGGGGCAGGT
44740
44799
GACGACGCATTGCAGTAACAGGTGGCTGTGTGCGTGCCGTGCCCACTTTCACCCCGTCCA
Q Q T V D N D V S V R V A R I E I P C I
CATGAATGAGGCGCTGAGGTCCCTGATCATGCCCACGGTGGGGCGGAGGTCGGAGATCTC
44800
44859
GTACTTACTCCGCGACTCCAGGGACTAGTACGGGTGCCACCCCGCCTCCAGCCTCTAGAG
M E S A S L D R I M G V T P R L D S I E
CAGCAGATCCCTGAGCGTCCCATTCTCCAAATTGTCGAGGATGTCCTCGTCCCTGGTAAA
44860
44919
GTCGTCTAGGGACTCGCAGGGTAAGAGGTTTAACAGCTCCTACAGGAGCAGGGACCATTT
L L D R L T G N E L N D L I D E D R T E
ATGGTGGCTGAAGGCTGGCCCGTTGTAGGCCAGGGTCTGGGCCACGTGCTGAAAGTCCAC
44920
44979
TACCACCGACTTCCGACCGGGCAACATCCGGTCCCAGACCCGGTGCACGACTTTCAGGTG
H H S E A P G N Y A L T Q A V H Q E D V
CCCGAGGCCGCACATGTGGGCATTGGTGCAGGTTGGGAGGAAAACGTAGTAAAAGATCTT
44980
45039
GGGCTCCGGCGTGTACACCCGTAACCACGTCCAACCCTCCTTTTGCATCATTTTCTAGAA
G L G C M H A N T C T P L E V Y Y E I K
ITCCAGCACATCCGCATGCCCCTCATCTACATAAGGGCCTAGGTGCAGACGGAAATCGTG
45040
45099
AAGGTCGTGTAGGCGTACGGGGAGTAGATGTATTCCCGGATCCACGTCTGCCTTTAGCAC
E L V Q A H G E D V Y P G L H L R E D H
GTCGTGGTCTCCGTTAACCCGGTAGCCGTACAAGGCCACAAATTGGGCAGCCATCTCATC
45100
45159
CAGCACCAGAGGCAATTGGGCCATCGGCATGTTCCGGTGTTTAACCCGTCGGTAGAGTAG

```
d        D  H  D  G  N  V  R  Y  G  Y  L  A  V  E  Q  A  A  M  E  D

         CATGTTTCCAACCCTCTCAATAAACTGGGGCGCGGCCAGGGTGTCAGCGTAAACCTCATT
45160    +---------+---------+---------+---------+---------+---------+--------- 45219
         GTACAAAGGTTGGGAGAGTTATTTGACCCCGCGCCGGTCCCACAGTCGCATTTGGAGTAA

d        M  N  G  V  R  E  I  E  Q  P  A  A  L  T  D  A  Y  V  E  N

         TCCGATAATAATCTGGGGGGCCCGGTCACTAACGGTGAGAAGATGGGTGAAAATGTCTGT
45220    +---------+---------+---------+---------+---------+---------+--------- 45279
         AGGCTATTATTAGACCCCCCGGGCCAGTGATTGCCACTCTTCTACCCACTTTTACAGACA

d        G  I  I  I  Q  P  A  R  D  S  V  T  L  L  H  T  E  I  D  T

         GTAGGCCACCGGGGGGAGCAGGTTAGGGTCCAGGAGAGCGCAGACATACTGACCCACGCT
45280    +---------+---------+---------+---------+---------+---------+--------- 45339
         CATCCGGTGGCCCCCCTCGTCCAATCCCAGGTCCTCTCGCGTCTGTATGACTGGGTGCGA

d        Y  A  V  P  P  L  L  N  P  D  L  L  A  C  V  Y  Q  G  V  S

         CTCATCCCCCACAACATCTGACCCGGCCAGGCGCATCAGGGCCTGCTCTAGGGCTATAAG
45340    ---------+---------+---------+---------+---------+---------+--------- 45399
         GAGTAGGGGGTGTTGTAGACTGGGCCGGTCCGCGTAGTCCCGGACGAGATCCCGATATTC

d        E  D  G  V  V  D  S  G  A  L  R  M  L  A  Q  E  L  A  I  L

         ITCCCCATAGATTTTTCTATACATGGAATAGGCCTCCTTGGAGATGGCGTTATTTCCCAG
45400    +---------+---------+---------+---------+---------+---------+--------- 45459
         AAGGGGTATCTAAAAAGATATGTACCTTATCCGGAGGAACCTCTACCGCAATAAAGGGTC

d        E  G  Y  I  K  R  Y  M  S  Y  A  E  K  S  I  A  N  N  G  L

         GTGGCGGCAGATGAAGTTGATCATGGAAAAGCTGTTCACAAAGGCAAGCCTCCCTGACCG
45460    +---------+---------+---------+---------+---------+---------+--------- 45519
         CACCGCCGTCTACTTGAACTAGTACCTTTTCGACAAGTGTTTCCGTTCGGAGGGACTGGC

d        H  R  C  I  E  K  I  M  S  S  S  N  V  E  A  L  R  G  S  R

         TTCCCAGTAGGTGTTGATGCACAGGGACACCAAAGGCACGTTCATGACAAACTTTTCCTC
45520    ---------+---------+---------+---------+---------+---------+--------- 45579
         AAGGGTCATCCACAACTACGTGTCCCTGTGGTTTCCGTGCAAGTACTGTTTGAAAAGGAG

d        E  M  Y  T  N  I  C  L  S  V  L  P  V  N  M  V  E  K  E  E

         AAACCCGTGGATCATAGCCTCGACTACGTAGAAGAAGGCTGGATAGGCAGTGTCATAGGC
45580    +---------+---------+---------+---------+---------+---------+--------- 45639
         TTTGGGCACCTAGTATCGGAGCTGATGCATCTTCTTCCGACCTATCCGTCACAGTATCCG

d        E  G  H  I  M  A  E  V  V  Y  E  E  A  P  Y  A  T  D  Y  A

         AGTATCCTGCACAGTCTCAATAACGGCCTGATCCACCACGTGGGCCAGAGATGTGGCGGT
45640    +---------+---------+---------+---------+---------+---------+--------- 45699
         TCATAGGACGTGTCAGAGTTATTGCCGGACTAGGTGGTGCACCCGGTCTCTACACCGCCA

d        I  D  Q  V  I  E  I  V  A  Q  D  V  V  H  A  L  S  T  A  T

         CTCAAACTGCTGCCCCCGGGCCTCTTGGAATGCAGCTGGGGCCAGGGGAGTCGGCAGGTT
45700    +---------+---------+---------+---------+---------+---------+--------- 45759
         GAGTTTGACGACGGGGGCCCGGAGAACCTTACGTCGACCCCGGTCCCCTCAGCCGTCCAA

d        E  F  Q  Q  G  R  A  E  Q  E  A  A  P  A  L  P  T  P  L  N

         ACCCACCATTAGCCGGTGCACAGCCCTGTGCCTGGCCCTCTCCCCGGCATCCCTGCCAAT
```


45760
45819
TGGGTGGTAATCGGCCACGTGTCGGGACACGGACCGGGAGAGGGGCCGTAGGGACGGTTA
G V M L R H V A R H R A R E G A D R G I
GTAAATATCATAAAGGGGGTGCAGCTCCAGCCGCAGCAGGTCATAATTGGACGGGTGGAG
45820
45879
CATTTATAGTATTTCCCCCACGTCGAGGTCGGCGTCGTCCAGTATTAACCTGCCCACCTC
Y I D Y L P H L E L R L L D Y N S P H L
GAAGTCTTCGGTGGGCAGCCCGCACTTGAGAGCTATATCTGTCACGGGGGCTGCATACTT
45880
45939
CTTCAGAAGCCACCCGTCGGGCGTGAACTCTCGATATAGACAGTGCCCCCGACGTATGAA
E D E T P L G C K L A I D I V P A A Y K
GTTATCATAGAACTCGTCCACAATAACAAGCACATTCATGTGATTGGGCCTCCTGTGTTG
45940
45999
CAATAGTATCTTGAGCAGGTGTTATTGTTCGTGTAAGTACACTAACCCGGAGGACACAAC
N D Y E E D V I V L V N M H N P R R H Q
CAGGGAGTAGGTCTCGCGCCTGTCTCGCGGGGCCGGGGCCGCGTTGAGGCTGTTTAGGGT
46000
46059
GTCCCTCATCCAGAGCGCGGACAGAGCGCCCCGGCCCCGGCGCAACTCCGACAAATCCCA
L S Y T E R R D R P A P A A N L S N L I
ATGGGCGGGTGTGTGGAGTCGGGGGTGACAGAGAACCTTGAGAGCATTCTGTAGGTTAAA
46060
46119
TACCCGCCCACACACCTCAGCCCCCACTGTCTCTTGGAACTCTCGTAAGACATCCAAATTT
H A P I H L R P H C L V K L A N Q L N E
CGCGAGGAGAAGGTTATTCTTGTTTACGATCCATGCCTCCACCGGTAGCTGCTGTGTGGG
46120
46179
GCGCTCCTCTTCCAATAAGAACAAATGCTAGGTACGGAGGTGGCCATCGACGACACACCC
A L L L N N K N V I W A E V P L Q Q T P
GTTGTCCAGCATTTTGATGGCGGCGGAGGTCGTGTACTTGGGATTGGGCATAAACAGGCC
46180
46239
CAACAGGTCGTAAAACTACCGCCGCCTCCAGCACATGAACCCTAACCCGTATTTGTCCGG
N D L M K I A A S T T Y K P N P H E L G
CACTGGGAAATAGTAGCTGTACTGCATTCTTCTGTTGAGGGGGTATGGGGACTGAGTGTC
46240
46299
GTGACCCTTTATCATCGACATGACGTAAGAAGACAACTCCCCCATACCCCTGACTCACAG
V P E Y Y S Y Q M R R N L P Y P S Q T Q
ATTGTACATCTTTTGCAGGCTTTCCACGGCCACCGCGTGGTTGCCCAGCTTGATGACGGC
46300
46359
TAACATGTAGAAAACGTCCGAAAGGTGCCGGTGGCGCACCAACGGGTCGAACTACTGCCG
M Y M K Q L S E V A V A H N G L K I V A
GGCTGAGATCGGCACCCGGGGCTGATCCTCGACCCCTGCGGCCACAGCCGGCAGGTCAGA
46360
46419
CCGACTCTAGCCGTGGGCCCCGACTAGGAGCTGGGGACGCCGGTGTCGGCCGTCCAGTCT


A S I P V R P Q D S V G A A V A P L D S
CTTGGTGCTTCCGGCTTTTTCCGGTGAGTCCACGATCCTAGCCATGAAATGCTCAAACGT
46420
46479
GAACCACGAAGGCCGAAAAAGGCCACTCAGGTGCTAGGATCGGTACTTTACGAGTTTGCA
K T S G A K E P S D V I R A M E H S E I
ACGCATCACGCGCCCGTAGCTCACGGCAGTGACCAGGTTCTCCCCCCGTACCACAAAAGA
46480
46539
TGCGTAGTGCGCGGGCATCGAGTGCCGTCACTGGTCCAAGAGGGGGGCATGGTGTTTTCT
R M V R G Y S V A I V L N E G R V V E S
AGCATAGCTCGAGGGCCCCATAATCTGGTTGTCGGCCTCCTCACCCAGGAAGGTCAAGAG
46540
46599
TCGTATCGAGCTCCCGGGGTATTAGACCAACAGCCGGAGGAGTGGGTCCTTCCAGTTCTC
A Y S S P G M I Q N D A E E G L E I L L
CTGGCGCAGAACGTTGTCGGTGACAATAAACACCCCCCCCACTGGCTCTCCCCCCTTGGC
46600
46659
GACCGCGTCTTGCAACAGCCACTGTTATTTGTGGGGGGGGTGACCGAGAGGGGGGAACCG
Q R L V N D T V I E V G G V P E G G K A
GGTCGTGTAGGTACTGACCCCCTTGAGCACGCTCTCCCCGGACACGGCCGCTACCATCTC
46660
46719
CCAGCACATCCATGACTGGGGGAACTCGTGCGAGAGGGGCCTGTGCCGGCGATGGTAGAG
I T Y T S V G K L V S E G S V A A V M E
AGAGAGACGGCTTCGCACGTACTGAGAAAACCCGGAGCCCATGTTCTCGGCCCGGTCCAG
46720
46779
TCTCTCTGCCGAAGCGTGCATGACTCTTTTGGGCCTCGGGTACAAGAGCCGGGCCAGGTC
S L R S R V Y Q S E G S G M N E A R D L
GAAGAAGGAGTGCTCCAGCAGATGCCTCTTGAACATGGCAATGAGGTCAGACTTGACAGT
46780
46839
CTTCTTCCTCACGAGGTCGTCTACGGAGAACTTGTACCGTTACTCCAGTCTGAACTGTCA
E E S H E L L H R K E M A I L D S K V I
CTTGGAGAACCCCCTCTCAGTGAAGGTGGGATCCGCCAGGGTCTGCAGGATAAACATGGG
46840
46899
GAACCTCTTGGGGGAGAGTCACTTCCACCCTAGGCGGTCCCAGACGTCCTATTTGTACCC
K S E G R E T E I P D A L T Q L I S M P
AGGGGCATGGCGAAGCTTCACACTCAGGACGGTGTTAATGAGGCCCCTCTCCAGGGCATC
46900
46959
TCCCCGTACCGCTTCGAAGTGTGAGTCCTGCCACAATTACTCCGGGGAGAGGTCCCGTAG
P A H R L K V S L V T N I L G R E L A D
GACCCCAAACTGTAGGGCCGAGGCCACGGTCTTGACAGCCCCCACGTACTCTGCGTACTC
46960
47019
CTGGGGTTTGACATCCCGGCTCCGGTGCCAGAACTGTCGGGGGTGCATGAGACGCATGAG
V G E I L R S A V T K V A G V Y E A I E
GACCGGGGTCTCGGGGATACTATGCAGGATCTCCAGATCCAGCATGGACAGTTCCATTTC

```
47020 +---------+---------+---------+---------+---------+---------+ 47079
      CTGGCCCCAGAGCCCCTATGATACGTCCTAGAGGTCTAGGTCGTACCTGTCAAGGTAAAG

d       V  P  T  E  P  I  S  H  L  I  E  L  D  L  M  S  L  E  M  E

      CGTACTAAATGTGGTGTTTGTGGCAAATTTTTGACCACAATGAATGTCCGCTGCTTGCTGGG
47080 +---------+---------+---------+---------+---------+---------+ 47139
      GCATGATTACACCACAAACACCGTTAAAAACTGGTGTTACTTACAGGCGACGAACGACCC

d       T  S  I  H  H  K  H  C  N  K  V  V  I  E  T  R  Q  K  S  P

      TCTCCTTCCGTCCCCGTGAGCAATGGTGGGGACGGAGATTCGAAATTGAATCTTGCCATC
47140 +---------+---------+---------+---------+---------+---------+ 47199
      AGAGGAAGGCAGGGGCACTCGTTACCACCCCTGCCTCTAAGCTTTAACTTAGAACGGTAG

d       R  R  G  D  G  H  A  I  I  P  V  S  I  R  E  Q  I  K  G  D

      CGTCATACGACTCAGGTCTTTGAATTCCGTGTTCACACAGGACACGGCCAGTGCCGTCTC
47200 +---------+---------+---------+---------+---------+---------+ 47259
      GCAGTATGCTGAGTCCAGAAACTTAAGGCACAAGTGTGTCCTGTGCCGGTCACGGCAGAG

d       T  M  R  S  L  D  K  E  E  I  N  V  C  S  V  A  L  A  T  E

      CAGGAAGCGAACATATTGGATGGCGTTCGTGTAGACCCCGAGTAGCACCTCAAACTTGAT
47260 +---------+---------+---------+---------+---------+---------+ 47319
      GTCCTTCGCTTGTATAACCTACCGCAAGCACATCTGGGGCTCATCGTGGAGTTTGAACTA

d       L  E  R  V  Y  Q  I  A  N  T  Y  V  G  L  L  V  E  E  K  I

      GCCCGCCTCTCTGGCATCGTTGCCCACCAGCAGGTCAAAGCTATGAAACAACCCCTCAGC
47320 +---------+---------+---------+---------+---------+---------+ 47379
      CGGGCGGAGAGACCGTAGGAACGGGTGGTCGTCCAGTTTCGATACTTTGTTGGGGAGTCG

d       G  A  E  R  A  D  K  G  V  L  L  D  E  S  H  E  L  G  E  A

      CGCTGACTGCCGCAGGTTCGAGAGCAGGTCGGCATCCACCGTCAGATAGGGGAAGGGTCT
47380 +---------+---------+---------+---------+---------+---------+ 47439
      GCGACTGACGGCGTCCAAGCTCTCGTCCAGCCGTAGGTGGCAGTCTATCCCCTTCCCAGA

d       A  S  Q  R  L  N  S  L  L  D  A  D  V  T  L  Y  P  E  P  R

      GTTTCCACACCCTCATTTGAGGCCATGACACAAGGTAAGAGGGAGATGGGGGGAGG
47440 +---------+---------+---------+---------+---------+---------- 47496
      CAAAAGGTGTGGGAGTAAACTCCGGTACTGTGTTCCATTCTCCCTCTACCCCCCTCC

d       N  E  V  G  E  N  S  A  M
```

Figure 29:
Expression of the ß-gal::p150 fusion proteins

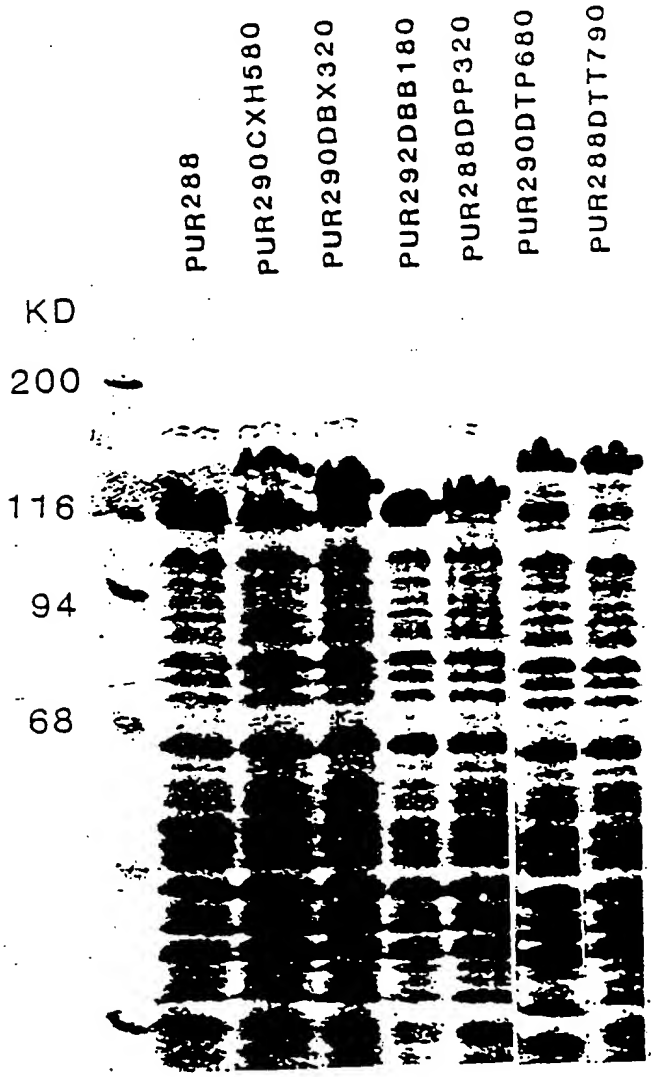
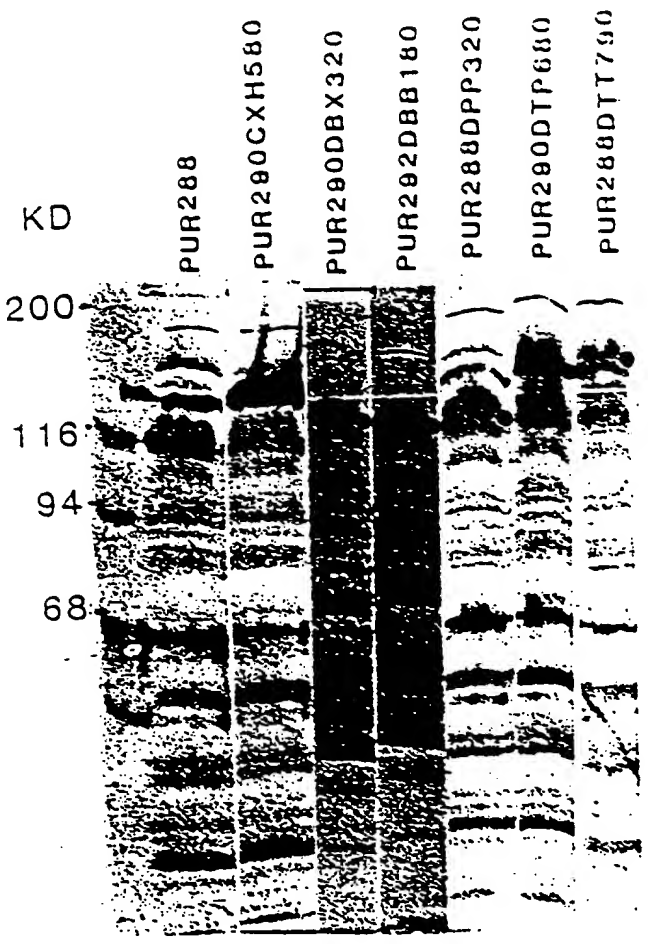


Figure 30:
Antigenicity of the ß-gal::p150 fusion proteins



European Patent Office

PARTIAL EUROPEAN SEARCH REPORT

which under Rule 45 of the European Patent Convention shall be considered, for the purposes of subsequent proceedings, as the European search report

Application number

EP 85110565.0

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.4)
P,X	EP - A2 - 0 151 079 (THE UNIVERSITY OF CHICAGO) * Claims 1-6 * --	1,12, 17,18, 24	C 12 N 15/00 C 12 P 19/34 C 12 N 7/00 C 12 N 1/20 C 12 N 1/16 C 12 N 5/00 C 12 P 21/02 C 07 K 13/00 A 61 K 39/245 G 01 N 33/53 C 07 K 17/00 //C 12 R 1:19 C 12 R 1:91
D,A	INTERNATIONAL JOURNAL OF CANCER, vol. 23, no. 6, June 15, 1979, Helsinki L.F. QUALTIERE et al. "Epstein-Barr Virus-Induced Membrane Antigens: Immunochemical Characterisation of Triton X-100 Solubilized Viral Membrane Antigens from EBV-Superinfected Raji Cells" pages 808-817 * Totality * ----	1,19	

TECHNICAL FIELDS SEARCHED (Int. Cl.4)

C 12 N
C 12 P
C 07 K
A 61 K
G 01 N

INCOMPLETE SEARCH

The Search Division considers that the present European patent application does not comply with the provisions of the European Patent Convention to such an extent that it is not possible to carry out a meaningful search into the state of the art on the basis of some of the claims.

Claims searched completely: -
Claims searched incompletely: -
Claims not searched: Claim 25
Reason for the limitation of the search:

(Article 52(4) EPC)

Place of search	Date of completion of the search	Examiner
VIENNA	19-11-1985	WOLF

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO Form 1503.1 03.82

0173254

European Patent Office

Application number:

85110565.0

DECLARATION PURSUANT TO RULE 28, PARAGRAPH 4, OF THE EUROPEAN PATENT CONVENTION

The applicant has informed the European Patent Office that, until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, the availability of the micro-organism(s) identified below, referred to in paragraph 3 of Rule 28 of the European Patent Convention, shall be effected only by the issue of a sample to an expert.

IDENTIFICATION OF THE MICRO-ORGANISMS

Accession numbers of the deposits:

pUCARG 680

DSM 3408